

nature

7 September 1978

Pakistan needs indigenous medicine

CHINA'S bare-foot doctor, little-educated, trained in his art for only six months, is a constant goad to Pakistan's health-care officials. Why can't Pakistan have similar people? For the conventional doctor with his Bachelor's degree (a five-year, full-time course in medical sciences) is highly expensive and scarce. And even if available, he is not mentally prepared to leave the attractions of the city for the rural areas. So he is ill-equipped to use his medical knowledge.

About three-quarters of Pakistan's people live in the remote countryside, and successive governments have committed themselves to provide a semblance of scientific rural health cover. But there is a largely untapped option in Pakistan, similar to the barefoot doctors of China: the indigenous medicine practitioners. Called the 'hakim', if they practise Greco-Arabic medicine ('Unani Tibb' locally), and 'ved' if they practise 'Ayurvedic', the ancient Indian system, they flourish in the countryside. Could they not be the raw material for a Pakistani bare-foot medical service?

Health care in Pakistan, as in almost the whole of South Asia, is a complex mixture of elements quite different from those in the west. For instance, there are in all about 12,000 conventional doctors in the country for a population of 70 million—hardly one doctor for every 6,000 persons. In practice these provide for only one-fifth of the population. Such a low ratio (for comparison, one doctor to 641 persons in the US, with 45 categories of health personnel to assist him) creates a health care vacuum that fills itself through hakims, veds, homoeopaths and even simple charlatans.

Some 5,000 qualified hakims with diplomas from Unani Tibb colleges already contribute to rural health care; but there are also another 40,000 unqualified hakims who have no formal education nor any medical training. These hakims have learnt just by sitting beside an experienced practitioner in his clinic. Nevertheless, they are recognised and registered practitioners of medicine in Pakistan.

As far as the norms of modern medical science are concerned, even the qualified hakims coming out of the institutions after a 4-year course have a motley education. Along with Greco-Arabic system—which largely works on empirical basis—the hakims are taught several subjects of modern medicine. So the 'kuliyaat-e-qanoon' of Avicenna is in the curriculum alongside books on modern physiology, anatomy and so on.

Dr Mazhar H. Shah, a Fellow of the Pakistan Academy of Sciences, writes in his recently-published book *the Unity and Integration of Medicine*, that "the teaching of modern subjects [to the hakims] is by people not really qualified to teach; and there are no facilities for laboratory diagnosis and hospital treatment. The result is that Tibb colleges

produce hakims who wear the stethoscope and prescribe both indigenous as well as imported drugs without any knowledge of basic science or pharmacology. This sort of teaching and practice is neither safe for the seriously ill, nor useful to the national health system".

The cost of modern medicine to provide health care to the entire population of the country, on the other hand, appears simply prohibitive and beyond the reach of a developing country. The medical bill for health services alone in Western countries far outweighs the total of the national budget of Pakistan.

So there is a growing realisation that some sort of integration is imperative between expensive modern medicine and the inexpensive indigenous system, which is largely based on cheap local herbs. Thus the hakim and the ved could be taught the essential principles of the indigenous system but also be adequately equipped with basic principles of modern medical science; and not in half-baked fashion as at present, but in a real professional way. A village doctor emerging from such a scheme could use his discretion to use either the local herb or, if available, the relevant Western drug.

The envisaged integrated system, however, may not find smooth sailing in Pakistan due to the many vested interests. The hakim and the ved is generally opposed to integration, as at present he tries to equate himself in social status with the fully qualified western doctor. So does the homoeopath (homoeopathy is a recognised system of medicine in the country). Strangely enough, a good cross-section of western doctors is also opposed to integration as they think that it would down-grade them if systems based on empirical principles are made to dovetail with scientific methods.

Political expediency under pressure from the practitioners of indigenous systems led the Government to set up Autonomous Boards for Tibb and Homoeopathy in 1965 so as to let them "train and practise their systems independently by supplementing such modern advances as the Board may, from time to time determine." However, the government also set up (in 1974) an Indigenous Medicines Commission with Dr Salimuzzaman Siddiqui, the chemist, as its chairman to investigate the possibility of integration of the three systems. The commission's report was never made public, but it is well-known that it had recommended integration to produce a health service structure on the lines detailed above. But a seal maybe put on the issue by the next civilian government-out of political expediency and to gain political advantage from the strong lobby of tens of thousands 'hakims'. But then that might be the end of integration and also of a practical, down-to-earth answer to rural health care in Pakistan. □



Carter juggles with fast breeder and natural gas

John Douglas discusses how President Carter has managed to enrage combatants on all sides of the fast breeder issue

PRESIDENT Carter has further entangled himself in the meshes of his energy bill — this time through an unlikely deal linking the price of natural gas with research on the fast breeder reactor.

Carter's proposals for price controls on natural gas had been bottled up in a Congressional conference committee for six months, and to free the gas bill for consideration by the full Senate, he needed one more Republican signature on the conference report. So he invited Senator James A. McClure of Idaho to the White House to discuss a compromise. Both parties deny actually making a 'deal', but McClure signed the conference report and an-

Schlesinger also committed the Administration to building the Safety Research Experiment Facility (SAREF), the Fuels and Materials Evaluation Facility (FUMEF) and the High Performance Fuel Laboratory (HPFL) — all of which the President had earlier opposed as part of his policy to shift from "early commercialisation to long-range technical development" of the breeder.

Finally came the matter of the Clinch River Breeder Reactor (CRBR), the costly and controversial demonstration project Carter has opposed since early in his election campaign. Schlesinger wrote that the project would be "discontinued" — without

own party on the issue to ignore the Republican votes Baker can influence. But more important, the compromise suggested to Baker that even if a large breeder reactor were finally authorised, the chances of building it at Clinch River, Tennessee, had been diminished.

Faced with a tough election campaign for his own Senate seat this fall, Baker reacted angrily to what he saw as a personal slight and a threat to a billion-dollar project in his home state. He announced that unless the status of the Clinch River project were cleared up he would no longer support the natural gas bill, which faces strong opposition from a small but determined group of Senators. Without his apparently vital support to cut off debate, under the rules of the Senate a minority of members might well be able to kill the bill by filibustering.

The issue of natural gas deregulation is by far the most complex and controversial section of the National Energy Act finally released by House and Senate conferees. The other three sections — dealing with energy conservation, conversion of power plants from oil to coal, and reform of utility rates — will probably pass easily, but none is likely to have the international impact of the natural gas bill, whose purpose is to lower American oil imports by substituting domestically produced gas.

The conference committee report is the latest salvo in a 39-year confrontation between producers and consumers over government control of gas prices. If enacted, the new law would eliminate the distinction between gas sold within a state and that sold between states. The two categories are now controlled differently. The bill provides that *newly discovered* gas in both categories would be allowed to escalate in price somewhat more rapidly than inflation until 1985, when all price controls would be removed.

The argument in favour of this loosening of control is that creation of a single gas market will make more gas available to interstate buyers, encouraging them to substitute gas for imported oil. At the same time, this argument goes, rising prices will encourage producers to step up exploration for more new gas. The Department of Energy estimates that by 1985, the equivalent of one million barrels per day of imported oil will have been replaced by domestic gas, representing a 10 percent reduction in total imports, at a \$5 billion annual savings.

Opponents of the gas bill argue that higher prices will only fuel inflation and that recent history indicates that

Sorry, for copyright reasons some images on this page may not be available online

White House meeting: President Carter, Senator McClure, Energy Secretary Schlesinger.

nounced that he had received assurances from the White House that increased funds for a large breeder reactor no longer faced a Presidential veto. Presumably, some of these reactor funds would find their way to the Idaho National Engineering Laboratory, which has long experience in the breeder field.

Details of the understanding reached on the breeder issue were confirmed in a letter from Energy Secretary James Schlesinger, who also attended the White House meeting. The Administration, he said, would no longer oppose development of a conceptual design for an "improved demonstration project" of the plutonium breeder reactor, although the size was not specified. To fund the entire breeder program \$513 million would be appropriated for fiscal year 1979, \$504 million for FY 1980, and \$520 million for FY 1981. (The Administration had originally wanted to cut back such funding to about \$400 million for the next fiscal year, but the House had already defeated that proposal.)

specifying whether that meant "cancelled" (the earlier wording of Administration policy) or merely "deferred." Some components of the CRBR would be completed and tested, and some of the preliminary licensing would be conducted. A final decision on whether to build this or another large breeder reactor would then be made by 1982.

At first the compromise looked as if it might work. The Administration had overcome a major hurdle facing an important piece of legislation by giving in on a point on which it had already suffered a partial defeat. But in the process of hammering out this compromise, the President and his advisers committed the sort of political blunder that has too often marked the Administration's dealings with Congress: they ignored the Senate Republican leader Howard H. Baker Jr of Tennessee.

Under any circumstances, Senator Baker's cooperation would have been important to the ultimate passage of the natural gas bill, for the President faces too much opposition within his

Camera Press

gas companies will not rush out to drill more wells just because prices are rising. A Chase Manhattan Bank study of the question concluded that "the legislation could actually result in some weakening of the dollar as foreign exchange markets come to perceive that the long-run supply of natural gas in the U.S. is as likely to be reduced as increased." And at least two senators have announced plans to oppose the bill because of their opposition to breeder reactors, now linked to it.

The Carter-McClure compromise is also likely to broaden the debate over conduct of the entire breeder reactor programme. Work is still not complete on the Fast Flux Test Facility, (FFTF) a non-power-generating breeder reactor

meant to be a forerunner of Clinch River. First authorised in 1967, FFTF was originally estimated to cost only \$87.5 million but has already passed the billion dollar mark. The CRBR has also experienced severe delays and cost overruns, and both industry and Congressional sources tell *Nature* that any further delay would probably result in a re-evaluation of the basic American breeder design.

During the debate over whether breeder reactors should be built at all, many technologists have reportedly swallowed their specific objections to the CRBR design in hopes of not exacerbating the issue. Specifically, they favour a reactor in which the radioactive molten sodium that passes through the core never leaves the re-

actor vessel and one in which steam for generating electricity is handled at a lower temperature and pressure than now called for. If the Clinch River project is deferred or cancelled in favour of producing a later—and probably larger—reactor, such arguments will probably become much more audible.

It is hard to recall a time when the US energy picture has appeared more cloudy. After he had calmed his initial anger, about the best Senator Baker could say for the recent compromise was "the water has been muddied". All that remains certain is that both the natural gas regulation and breeder reactor issues will take on new dimensions when Congress returns from its present Labour Day recess. □

Boycott of Soviet contacts is for individuals, says NAS

RECENT 'human rights' trials of scientists in the Soviet Union resulted in the immediate cancellation of several major visits by US scientists to the Soviet Union—and these in turn have provoked a biting reaction from the Soviet media—a sure sign that the threat of a severance of scientific relations was a real one. "Future historians will be amazed" wrote the *Literaturnaya Gazeta*, that such reprisals could be contemplated as a means of making the USSR give up "the principles which are sacred to it."

This reaction would appear to substantiate the case for boycott, a case most ably put by Valentin Turchin (*Nature* 273, 256–257; 1978). Not all scientists would support this view, feeling that continuation of contacts makes it possible to express one's concern with the fate of, say, Orlov and Shcharanskii, at least on the personal level, to Soviet colleagues encountered at international conferences.

Thus last month, when the International Congress of Mathematicians was held in the "human rights" city of Helsinki, right on the Soviets' doorstep (subsidised excursions to Leningrad were laid on as a side attraction), several Soviet mathematicians failed to arrive to deliver their papers. One of these people, a Dr Margulis, was scheduled to receive a medal. The ceremony was carried out in his absence, and the formal "presentation" of his work was greeted by a standing ovation, pinpointing in a telling manner the persistent Soviet practice of denying visas to scientists invited to international conferences.

Many would see the stricter insistence on the norms of scientific life—including proper representation at con-

ferences, a more fitting form of protest than an all-out or selective boycott.

Respecting the diversity of views on boycott, the US National Academy of Sciences recently put out a statement calling on world scientists to urge the release of Orlov, Shcharanskii and Sergei Kovalev, a biologist now halfway through a sentence of seven years in a labour camp. The Human Rights Committee of the Academy said:

"These scientists, along with fourteen others in other parts of the world, are of particular concern to this committee, for they represent numbers of colleagues who believe that freedom of intellectual inquiry cannot be divided into two parts—one for the natural world in which we live and one for the society which nurtures us—and who have been imprisoned for acting in accordance with their beliefs".

Nevertheless, the Academy stressed that it does not, as a body, endorse boycotts. "Each American scientist", it concludes, "contemplating a visit to the USSR (or asked to host a Soviet scientist in the US) must determine his or her own course of action."

Interestingly, this decision follows an opposite course to that advocated by a group of French physicists last July, who advocated the severance of all official scientific relations—including conferences—while maintaining personal contacts with Soviet colleagues, making the latter the "occasion of expressing our indignation".

The NAS acceptance of individual choice of protest also underlies the concluding paragraph of a moving appeal which recently reached the West from Kovalev's son, Ivan. Addressing the intending participants of the Fourteenth International Congress of

Genetics in Moscow, he outlined the history of his father's case, and gave a vivid and horrifying picture of prison conditions, he asked for the participants' support, that they would express their attitude towards the imprisonment of their colleague "by means appropriate to you". The appeal has since been circulated by campaigners for boycott, but it is noteworthy that neither Amnesty International, who distributed the letter, nor Ivan Kovalev himself, specifically ask for this.

Nevertheless, the idea of at least a selective boycott does seem to be taking hold in the USA. According to the NAS, its supporters "include scientists who pioneered in the earliest Soviet/US exchanges, seeking to build bridges of common scientific endeavour across the chasm of the cold war. They also include others who have seen themselves as steadfast in resisting the politicisation of science. People have reached their decisions in varied ways: sadness, rather than anger has been the most common emotion".

It is because the response of US scientists is so "individualistic", concludes the Academy, that continued Soviet-US scientific relations are in peril. "Scientific exchange programmes" it says "can be negotiated and organised, but individual participation cannot be commanded. There has been no institutional instruction or decision making, no rush to judgment, and no stampede to boycott. Rather there is a tide of spontaneous response running deep, and it will not be easily reversed in the absence of some judicious and humanitarian actions by Soviet authorities".

Vera Rich

Soyuz is not a trolley-bus, but Comecon flights continue

SIGMUND Jähn, the first East German cosmonaut, together with his Soviet flight-partner Valerii Bykovskii, returned to Earth last Sunday, after a week aboard the Salyut-6 orbiting station. Jähn, the third non-Soviet cosmonaut to visit Salyut, completes the first phase of "international" flights; further launches, with the participation of cosmonauts now in training from other Comecon countries are scheduled for next year.

The manned "Interkosmos" programme, in which the Soviet Union puts into orbit crew members from her Comecon allies did not at first meet with the approval of the entire Soviet scientific establishment. Academician Anatolii Aleksandrov, President of the Soviet Academy of Sciences, was reported to be against any such non-Soviet participation ("Soyuz is not a trolley-bus!"). The programme appears to have been inaugurated partly as a visible expression of fraternal solidarity, and, more practically, as a recompense for the contributions made to the Soviet space programme by the Comecon allies. (It is notable that the first three countries to be honoured by a flight were Czechoslovakia, Poland, and East Germany, whose contribution has been particularly great).

As regards Jähn's particular contribution to the joint mission, one experiment, at least, codenamed "Rech" makes use of his linguistic background. During all communication sessions, he had to repeat "Zwei hundert sechs und zwanzig", so that the parameters of his speech—tone, volume, rate, etc., could be monitored and related to his physical and psychological reaction to weightlessness. (The significance of the number 226 is not clear—nor why the experiment could not have been performed using a Russian repeating "Dvesti dvadtsat' shest'!")

Reaction to weightlessness was the underlying theme of most of the East German experiments, including tissue culture and bacteria metabolism experiments, and an investigation of the cross-linking of micro-organisms and organic polymers. This last used "a Soviet apparatus modified by GDR specialists" (essentially a complex of polythene bags).

And, as in previous "international" flights, the "Splav" furnace was used to produce semiconductor compounds—in this case lead-tellurium and bismuth-antimony compounds, the experimental capsule being renamed for the occasion "Berolina". **Vera Rich**

Sorry, for copyright reasons some images on this page may not be available online

Taj Mahal out of danger?

THE Indian Government is considering a report from a committee of experts, headed by the chairman of the Indian Petrochemicals Corporation, Dr S. Varadarajan, on how pollution from an oil refinery being built near Agra will affect the Taj Mahal. The committee began its investigations four years ago after fears in scientific circles and in the press that sulphur dioxide waste from the refinery at Mathura 146 km south of New Delhi, would damage the white marble of the Taj Mahal and red sandstone of other historical monuments at Agra, 40 km farther southeast.

The committee's studies revealed that the sulphur dioxide content of the air in Agra is already 15 to 20 micrograms/m³ because of two power plants, numerous small industries and a railway shunting yard there. The refinery, according to the report, would only add a further 1 to 2 micrograms/m³ in the long term, although short-term peak concentrations of 65 micrograms/m³ might be expected under the worst meteorological conditions in winter.

Meanwhile the Government also engaged an Italian firm, Techneco, which specialised in the damaging effects of pollution on marble, to study the problem. The firm in its report ruled out any potential threat to the Taj from the refinery. Another study by the Indian Meteorological Department (IMD) on the expected ground level concentrations of sulphur dioxide in the Agra region because of the refinery also sought to allay the apprehensions. None of the two reports have been made public so far.

However, some experts and technical agencies have expressed doubts about the validity of the findings of Techneco and IMD which they say are based on wind data not of Agra but of Delhi. They alleged that the mathematical models, which were developed abroad, and the constants used were not appropriate for the local weather conditions. The Archaeological Survey of India, which is responsible for the maintenance and conservation of the monuments, is of the view that the emissions from the refinery would certainly discolour and wear away the white marble of the Taj.

Further critics say the Varadarajan committee has avoided the main issue: the report does not say in unequivocal terms whether the threat does or does not exist. They point out that the terms of reference did not include the possibility of shifting the refinery to another site away from Agra. (The Minister of Petroleum and Chemicals, Mr H. N. Bahuguna, had reportedly given the assurance last year that he would reconsider the moving of the refinery to Etawah, 100 km east of Agra, if the expert committee definitely said there was potential danger to the Taj.)

Nevertheless the committee recommends some control measures: that long-term emission of sulphur dioxide from the refinery should be limited to 1 to 2 micrograms/m³ and that three monitoring stations should be set up between the refinery and Agra to keep a constant check on this. The committee also puts forward plans to reduce the overall sulphur dioxide

content in Agra by closing the power plants, switching from coal-fired locomotives to diesel in the shunting yard, moving out some of the present industries and banning from the area any new ones which might add to pollution. It also wants to set up a body to monitor the air at Agra and advise industries on how to keep within acceptable pollution limits.

The Indian Oil Corporation, which has commissioned the refinery, has also taken steps to minimise pollution. Low sulphur fuel oil and gases will be used in the furnaces, and a plant will be installed for the removal and recovery of sulphur. The height of the emission stacks will be increased from 40 metres to 80 metres to ensure better dispersal, and fixed and mobile stations will monitor the concentration of sulphur

dioxide at ground level at various distances from the refinery. These measures will, according to the corporation, limit sulphur dioxide emission from the 6 million tonne refinery to 1 tonne per hour. Installing a planned flue gas desulphurisation unit could reduce this to 0.4 tonne per hour, the Corporation believes.

Corrective action is envisaged if the emissions rise above the safety level. But, critics ask, who will give directions to the refinery each time the monitoring stations give warning that the pollution has risen above the danger level? Moreover, they point out, the recommendations are not mandatory. What is the guarantee that the measures suggested by the expert committee would be fully implemented? And even if all the safeguards are taken, a certain amount of

residual pollution will cause damage to the monuments over a long period.

The latest among the public figures to come out openly against the refinery is the noted ornithologist Dr Salim Ali. In a public statement he called for the shifting of the refinery as "there is a controversy among experts about its effects on the Taj". Dr Ali also spoke about the threat to the famous Bharatpur bird sanctuary near Agra because of refinery gases.

While the Government is yet to take a decision on the Varadarajan Committee report, more than 80 million rupees is reported to have already been spent on the refinery project. There is growing apprehension among critics that it may soon be too late to turn back. □

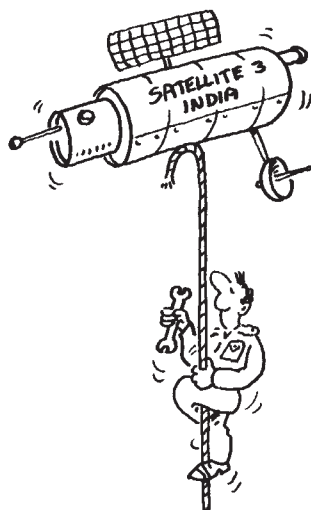
India plans to invest in tethered balloon technology

WITH the successful ascent of a 270-cubic metre tethered balloon at the Space Applications Centre (SAC), Ahmedabad, on 19 May this year, India has joined the select band of countries like the United States, Britain and France which have developed tethered balloon technology.

According to the Electronics Commission's Information, Planning and Analysis Group, India has all the expertise and indigenous materials to develop this technology on a large scale. A panel on tethered balloon technology, set up by the commission to work out the best plan given available resources and the work in progress in the country, has recommended developing balloons of all sizes—small, medium and big. The panel has also stressed the need for making possible users aware of the potential of balloons.

Research in this technology began in India when tethered balloons were seen as a substitute for the ATS-6 satellite for the continuation of the SITE programme. On the initiative of the Electronics Commission, research was taken up at SAC in April 1974. Apart from the many government and private institutes that helped in designing and testing the balloons, collaboration was sought with the Tata Institute of Fundamental Research, where work was already in progress on plastic balloons.

Two years ago, Tata launched a similar balloon to the present one from its National Balloon Facility at Hyderabad. The present balloon is of medium size, has a payload of 50 kg and can float to an altitude of 1.5 km. Except for two fireproof motors in its pressure regulation system, its components are all Indian made. It carried telemetry and telecommand equipment and batteries.



There is no doubt that tethered balloons are the poor nations' substitute for satellites. Once their technology is developed, they can be used for TV and FM broadcasting, remote sensing, ecology, oceanography, meteorology, coastline management and pollution monitoring. Because they are tethered, height can easily be varied for the study of atmospheric phenomena close to earth. A study carried out by IPAG suggests using balloons in offshore oil exploration as a temporary communication link with offshore platforms.

For communication and TV relays over hilly areas, where building earth-

bound installations is difficult and expensive, tethered balloons can also be used. The panel has said that two balloons at an altitude of 30,000 metres, one at Shillong and another at Dibrugarh, could be easily and cheaply used for a service covering most of the north-eastern hills of India.

The panel also advocates the immediate development of low technology tethered balloons—of low payload and short flight duration. For instance, a tethered balloon with a 10 kg payload, parking at an altitude of 1 km for 24 hours, could be useful to the Meteorological Department and the Institute of Oceanography. Such balloons could also be of use in the MONEX programme scheduled for the 1979 monsoon in the country. The panel has also asked for a study of hazards to the balloons, such as lightning and ultraviolet rays.

Once the technology is fully developed, it is recommended that a national agency should take over to tailor the balloons to the needs of users. How cost effective a tethered balloon can be will be apparent from one example. For a TV broadcast operation for one per cent coverage of India, the total capital investment is 40 million rupees with an annual recurring cost 4.2 million rupees. The same costs for a tethered balloon would be 6.5 million and 1.7 million rupees.

Dilip M. Salwi



Correction

In 'India prepares to launch Earth resources satellite' (10 August, page 526) it was incorrectly stated that the SEO satellite will be launched from the SHAR range in India. The satellite will be launched in the Soviet Union.

India on stand-by for a plague of locusts

INDIA is standing by for a locust invasion on its borders with Pakistan. Five wireless-linked locust control centres have been established in the border states, and chemicals, men and machines rushed in as locust activity increases. An agreement has been signed between the two countries for a joint strategy to combat the menace. Field officers on either side of the border are meeting regularly to exchange information on locust swarms, egg-laying and emergence of hoppers.

If the locust threat materialises, it will be the second invasion this year. It was in mid-June that the first swarmlets of locusts invaded India. A rare change in wind pattern over the Arabian Sea brought them to Banaskantha and adjoining districts of Gujrat from their original breeding grounds in Africa. The Plant Protection Organisation of the Agriculture

Ministry swung into action and neighbouring states alerted. Benzene hexachloride dust, dusting machines and handpump sprayers were rushed in.

Swarmlets swept in across the India-Pakistan border, and spread to Rajasthan, Haryana and Punjab. They came in small and thin waves. A swarmlet was hardly larger than 3 sq km in area—though a swarm as large as 27 sq km was also spotted in Rajasthan.

India's last locust plague, 16 years ago, destroyed half a billion rupees of standing crop. Except for spoiling leaves and some standing cotton crop in Haryana, the recent invasion did almost no damage—thanks to timely action by the Agricultural Departments of Punjab and Haryana.

Control rooms were set up and anti-locust squads formed. Villagers were trained to report any sign of locust activity. Areas where the locust swarms

had laid eggs were ploughed or sprayed with chemicals. The conventional methods of flame-throwing and digging trenches were also used in some villages, and eventually, all the swarmlets were either killed or paralysed, except one which managed to fly back to Pakistan across the Sutlej river.

The current locust concentration is in the largely desert Saurashtra-Kutch region of Gujrat. At this time of the year, heavy rains turn it into a marshland and a breeding and egg-laying ground for the locusts. Locust swarms have also laid eggs in a large number of villages in Rajasthan.

A total of 100 vehicles, five planes, two helicopters, 100 power dusters and 5,000 hand dusters have already been rushed to these border areas. Border Security forces and the Air Force are also on hand if needed.

Dilip M. Salwi

Vietnam still looking for foreign aid

VIETNAM'S appeal for more development aid seems to be having a mixed reception. Many countries and international agencies have signed technical and scientific aid agreements with the Vietnamese, but some now have reservations about starting new programmes. The United States refuses to consider any commitment to Vietnam until diplomatic relations are restored.

International organisations provide a substantial proportion of Vietnam's current aid revenue, and a number have commitments within the country. The World Food Programme, for example, has provided US \$89 million since April 1975 for emergency assistance, irrigation work, resettlement programmes in the 'new economic zones' and afforestation programmes. The United Nations Children's Fund is giving \$73 million from 1976-1979 to provide basic health care and primary education and to improve drinking water supplies. In addition the U.N. Family Planning Association (UNFPA) is already operating family planning programmes and will manufacture contraceptives; and the World Health Organisation (WHO) is, to take one example of its activities, training health care personnel.

Financial institutions such as the World Bank and the Asian Development Bank are also to make funds available for projects to improve irrigation, rehabilitate light industry, explore oil resources, and develop telecommunications.

UNESCO and the International Atomic Energy Association (IAEA)

are two of the more scientifically-orientated agencies with an interest in Vietnam. UNESCO is helping to equip the Nha Trang Centre of Oceanography and to set up the Scientific Information and Documentation Centre. (Because of its concern in the field of the restoration of historical monuments, it is also being asked for assistance in restoring the former imperial capital of Hue.) The IAEA feels it could help with the application of isotope and radiation techniques to agriculture and ground water surveys—both given top priority in the country's development programme.

Vietnam is no longer at the top of the UN list of those in need of development assistance—that position is now taken by Angola and Mozambique. But interest in Vietnam has not waned. The report back in March of the Special Commission on Assistance to Vietnam was very well attended. The Soviet Union, China, and the nations of Eastern Europe have long supplied aid to Vietnam and more recently Sweden, Norway and Holland have also established commitments in the country. They are maintaining a keen interest in the redevelopment of the country and are involved in several specific projects.

Vietnam has given science and technology high priority, but most aid agreements favour technology. Those concerned more with science, and involving Western nations, are limited to India, France, West Germany, Canada, Switzerland and the Philippines. These agreements are mainly to provide

advanced training for Vietnamese students.

Lack of technical equipment presents great problems for scientists in Vietnam. This means that research targets are not always met and—for biologists in particular—research programmes are sometimes postponed. The United Nations mission to Vietnam highlighted the needs of Vietnam's scientists, and the UN officials involved with the mission do not hide their disappointment at the apparent disinterest shown by countries supplying scientific aid. They admit, however, that they are not privy to the terms of every bilateral aid agreement and that some countries may be assisting by supplying scientists with vital equipment.

Britain, however, has no immediate plans to supply scientific aid. It is concerned with human rights in Vietnam. The Ministry of Overseas Development (ODM) and the Foreign and Commonwealth Office say that aid already committed to Vietnam will not be affected, but that no new agreements will be made until the government is reassured that human rights are not being violated.

But Britain is not alone. Sweden and Holland have expressed similar concern. They are worried about press reports of refugees leaving Vietnam; of the measures taken to restrict private enterprise in the south of the country; and of the means used to persuade people to move into the new economic zones. The British Government has also shown its concern about human rights in other countries and has

IN Britain today only about 400,000 people work full time on the land. Of these 160,000 are hired workers, 200,000 are farmers (tenants, land-owners and managers) and the rest are members of the farmers' families. Thus Britain, with a total population of 56 millions, has a smaller proportion of its citizens engaged in agriculture than has any other country, be it industrialised or what is now called, optimistically or euphemistically, "developing".

Yet notwithstanding this tiny labour force, food production in Britain is greater than ever before. Were we content with the diet enjoyed by all but the greediest and most prosperous individuals a hundred years ago, we should already be more than self supporting in all foods except things like tea and coffee which do not flourish in our temperate climate. We could survive a seige better than in the last war, when we had a smaller population to feed and nearly three times as many people working on our farms.

Farmers are, however, worried, because the small number of workers means that the industry has little political power. There is probably no single parliamentary constituency in even the most rural part of Britain where the votes of farmers, farm workers and their families would substantially affect the chances of election or rejection of a Member of Parliament.

Thus my own constituency of Huntingdonshire is perhaps the one which is most dependent on agriculture. At the last election the sitting member, Sir David Renton (Conservative) had a majority of nearly ten thousand votes. In this area there are 1,391 farmers, smallholders and market gardeners, and about the same number of employed farm workers. A few of these are below voting age, quite a number are single, and the total strength of the voting force is probably under five thousand. We do not know exactly how these

Farming lobby



KENNETH MELLANBY

people voted in the past, or how they will vote in the future. Large scale farmers are mostly traditionally Conservative, the workers trade union is to the left of the Labour Party, many smallholders are Liberal in their sympathies, so the farming vote is likely to be split. But even if 100% turned out for one candidate, the total would be less than half the present member's majority.

It is partly because their political power is so reduced that British farmers are making greater efforts to develop public interest and sympathy towards their problems. Many thousands of the general public attended the recent Royal Agricultural Show, where various entertainments were provided as well as displays of traditional farming activities. Up and down the country smaller county and district shows set out to cater for the non-farming public, giving them an enjoyable day out and a painless education at the same time. The town dweller at least learns that milk does not originate in bottles and tins, and that their daily bread comes from an arable crop which has its own pest and disease problems.

A recent development has been the organisation of "farm trails". On a Sunday afternoon in late June I went

along to a large arable farm in the fens near to my home. A conducted tour started off every hour. I joined about 150 miscellaneous men, women and children, from town and country. We mounted trailers to be drawn by enormous tractors around the fifteen hundred acre farm. The arrangements went smoothly, though the farm roads were bumpy, and one tractor (alas, the only British made model) broke down and its load had to crowd onto the four other trailers. We covered several miles on this totally arable farm. We saw fields of wheat, sugar beet, potatoes and green beans. At strategic points we stopped to be addressed by our guide, who explained what each crop was, and the main problems it presented. We were told why they kept no farm animals. We saw immense modern machines, for instance one to harvest beans (for freezing) which cost over £50,000 and was only used for about ten days in each year.

The crops, though not yet ready for harvest, were superb. It was obvious that yields would be greatly above the national average. Those who criticise large farms for being unproductive, in yield per acre, should have been there. Hardly a weed remained in the crops, though odd corners of fields were left uncultivated or planted with native trees and shrubs to encourage wildlife. Even a few hedges remained. This may be the "agribusiness" which many of my conservationist friends so criticise, but we could see how good farmers maintain and improve the fertility of their soil, and even care for the countryside and for the conservation of wildlife. In the depression in the nineteen thirties this farm was derelict, no one was prepared to work it even at the lowest rent. It is now producing enough food to feed all the population of a small town, and its yields continue to grow year after year. The demonstration was evidently an excellent exercise in public relations.

curtailed aid programmes to Uganda and Bolivia.

On the grounds of development alone, Vietnam is regarded by the US as well as Britain as being entitled to substantial aid. But there is little chance of the US supplying official aid in the near future. 1978 is an election year, and officials say that aid for Vietnam is too controversial an issue at present.

Vietnam and the US have not established diplomatic relations; both governments insist on certain preconditions before normal relations can be

restored. Senate officials have been assured, however, that unofficial links between scientists and institutions will not be opposed. Some US scientists are not convinced of the interest of Vietnamese scientists in such exchanges; offers of funding for study in the US have not aroused much response, and several scientists who visited Vietnam in the hope of studying the effects of herbicide spraying found the Vietnamese unhelpful. Such an attitude, however, is not without its political overtones (see *Nature* 271, 597; 1978). For this reason, some

scientists consider that it would be rash to prejudice the Vietnamese on their sensitivity over this issue.

US and UK attitudes to aid to Vietnam are unlikely to have disastrous consequences for Vietnam's current development programme. There could be problems, however, if more countries adopt similar policies. In this case Vietnam's scientists, desperately struggling to overcome the ravages of war on their research programmes would be some of the first to suffer. It is political decisions which will determine the outcome. **Alastair Hay**

correspondence

Nitrates and nitrites cannot be eliminated

SIR,—The report that nitrite has induced cancer in animals is not surprising as it has long been known to be mutagenic. The benefits from the use of nitrite as a food additive may outweigh the hazards that it presents as a mutagen and carcinogen. The amounts of nitrite in food either 'naturally' or as additives should obviously be kept to the minimum effective levels. Small amounts of nitrites and nitrates are formed by the electric discharge of lighting and it is difficult to avoid this contaminating drinking water.

Nitrate is, however, formed by the decomposition of organic matter and indeed the nitrate used for gunpowder was made in this way until the deposits of Chilean saltpetre became available in the nineteenth century. In conditions of poor sanitation the nitrate formed by "decomposition of organic matter" including faecal residues will be a pollutant. In dry regions this is likely to be present in air as a fine powder and in moist areas as a water contaminant. Improvement in hygiene could therefore remove or reduce a carcinogenic hazard. Nitrite is, however, formed in the intestine by bacterial action (Tannenbaum, S. R., Fett, D., Young, V. R., Land, P. D. & Bruce, W. R. *Science* 200, 1487; 1978) and it is difficult to see how this could be prevented.

Nitrates are converted to nitrites in the body. Nitrites could be carcinogenic directly by reacting with the purine residues of nucleic acid. They can also act by reaction with secondary or tertiary amines to form nitrosamines which may be metabolised by tissue enzyme to alkylating agents.

The new findings increase the correlation between mutagenicity and carcinogenicity and indicate that exposure to nitrates and nitrites should be reduced although they cannot be completely removed.

Yours faithfully,

E. BOYLAND

The London School of Hygiene
and Tropical Medicine

Foxing fungus

SIR,—The interesting letter from Maynell and Newsam (3 August, page 466), confirming the suggestion that foxed paper has suffered from fungal attack, raises the question "Why was this form of deterioration called *foxing*?" The dictionary explanation is that the word refers to the brownish colour. If that were all, rusting would have been a more suitable word. I have often wondered whether it is merely a coincidence that several harmful effects, now known to be caused by fungal growth, are verbally related to the fox. For example: fox-eil, now called alopecia from the Greek for a fox; wine fermented by a contaminated yeast culture; and beer that has 'gone off'.

Robert Graves argued somewhere

that fox in the Bible is often a mistranslation of *mushroom*. For example: "Take us the foxes, the little foxes, that spoil the vines" (*Song of Solomon* 2 15). He explained the extraordinary story (*Judges* 15 4,5) of Samson tying firebrands to the tails of pairs of foxes, and sending them off to burn the Philistine's crops, thus: "he inflamed his soldiers by feeding them on mushrooms (? *Amanita muscaria*) and sent them in pairs to destroy the crops".

I wonder whether there could have been a similar confusion in other languages. The metaphorical use of fox to describe colour or cunning is understandable. The uses of the word that are discussed here predate recognition of the fungal category. Nevertheless, they all have some connection with observable growth. We get fox from a Teuton rather than a Latin root. Could there have been a confusion between the German *wuchs* and *fuchs*. In English, the f becomes a v in vixen. It would be interesting to know whether any scholarly person has thought about the matter. It's not my subject.

Yours faithfully,

N. W. PIRIE

Harpden, UK

Compulsory boycott

SIR,—Unlike many other people, I never intended to boycott the Fourteenth International Congress of Genetics of Moscow, August 1978. I was very anxious to meet our Russian colleagues. Five communications from members of my department were accepted by the conference organising committee. Two of their authors had even been invited to act as co-chairman for some of the congress sessions.

But the Soviet authorities refused to grant me an entry visa, without giving any motive. To show their solidarity, my collaborators decided to cancel their participation in the congress.

In answer to my protestations made by telegram, I received (on 18 and 19 August) two calls from the organising committee, assuring me that there had been a technical mistake which would be rectified promptly. I was asked to wait for further information. I was still waiting, when the congress was almost at an end.

Who has boycotted the congress? Not I! But will my Russian friends hear of it?

Yours faithfully,

ANTOINE ELENS

Namur, Belgium

Making mincemeat of our common language

SIR,—An American visiting Britain, and wishing to cook a hamburger (whatever its load of carcinogens) is in for a "astronomic surprise if he acts on the information given by my friend Tom Jukes

(27 July, page 307). If, in an attempt to purchase ground beef he asks for mincemeat, he will receive "a mixture of currants, raisins, sugar, suet etc." (see the Oxford Dictionary), the stuff we British use at Christmas to make the traditional mince pies. If he is lucky, there may even be some brandy in the mincemeat; a hundred years ago our more virile ancestors included steak, but fruit and sweet substances always predominated.

The English synonym for ground beef is simply mince. The shopkeeper would understand what was needed if asked for minced beef.

Yours faithfully,

KENNETH MELLANBY

Monks Wood Experimental Station,
Abbots Ripton, Huntingdon.

Marsh mallow grows in Britain

SIR,—Rosemary Angel, in her review (4 May page 79) of *Major Medicinal Plants: Botany, Culture, and Uses*, takes the author, Julia F. Morton, to task for saying that marsh mallow is grown in gardens in Great Britain. "No, I am afraid it is not," she says. Without wishing to be picayune, I feel I should say that it is. I have grown marsh mallow in two successive gardens, and it is one of the most beautiful plants I know. It is growing in my garden now. And I am well aware of the difference between common mallow, musk mallow, marsh mallow, and the others.

Yours faithfully,

ROBERT K. G. TEMPLE

Bridgwater, Somerset, UK

Sex and cancer

SIR,—In his review of A. C. Braun's book, *The Story of Cancer: On its Nature, Causes and Control*, Richard Peto says, "... many, and perhaps most, cancers are caused by certain sexual habits, smoking habits, and gross aspects of diet rather than by contaminants."

For God's sake, Peto, which habits? Are thousands, perhaps millions of human beings to perish through your reticence (or bashfulness)?

And if—as can hardly be doubted—the Regius Professor is correct, can he tell us which sexual habits of plants cause their cancers? Or which smoking habits?

Yours in fear and trembling,

RALPH CHERNOFF

Los Angeles, CA, USA

news and views

Thin films: present and future

from Robert M. Hill

TWENTY-FIVE years ago, as part of an undergraduate course, I was introduced to the world of thin films. In an evacuated bell-jar a sample of aluminium wire was heated up by a tungsten filament, the aluminium melted then vaporised and condensed on the cold surfaces of the bell-jar and on a microscope slide mounted in a holder. I think the rest of the experiment was to measure the optical transmittance and reflectance of the semi-transparent metal film. The feature that caught my imagination was the extreme thinness of the metallic deposit, somewhat less than 30 atom layers. For the first time atomic dimensions were realisable, structures could be seen to grow and one could visualise the packing of atom layer on atom layer and investigate the physical effects of size down to angstrom dimensions.

The same period of 25 years has seen an amazing growth in the technology of thin films. New methods of deposition, or rather more correctly, improved techniques based on the development of old ideas have become available, a new range of instruments for investigating the structure and properties of films have been developed and there have been major steps in our understanding of the delicate processes of nucleation and growth, and of the effects of structure on the physical properties of films. A major step in the recognition of thin film science was the establishment, in 1966, of a journal, *Thin Solid Films*, specifically devoted to the publication of scientific papers on thin films and their properties. The journal has now brought out its 50th volume (gone are the days when volumes and years kept pace together) and has taken the opportunity of asking the editorial board and contributors to present their personal views of the past and of the future. Scientists, as we all know, are notoriously reticent

about forecasting the long term future, unless one manages to be party to a group relaxing together after a long day listening to interminable papers at a conference or summer school, and there is little crystal ball gazing in this anniversary issue but there is an excellent set of individual views of the past, and particularly the present, state of the field.

The methods of preparation of thin films are many. In this volume alone are descriptions of deposition by vacuum evaporation, chemical vapour deposition, ion beam methods, gaseous adsorption, photodeposition, plasma polymerisation, epitaxial overgrowths and the stacking of monomolecular layers. In principle it seems that any material, element, alloy or compound, can be grown in the form of a thin film. Indeed many alloys can be deposited in metastable forms that are not known in the bulk. The most perfect layers are those grown by epitaxy on single crystal substrates, a technique widely used in the preparation of electronic devices where the interface between film and substrate, or between two different films, of good crystal structure has to be sharp and well defined. Compound formation can be aided by vapour deposition in which chemical reactions are allowed to take place either in the vapour phase or at the growing interface. Much of the recent work on the electrical properties of amorphous semiconductors was carried out on thin film specimens in which the normal crystallisation processes were inhibited by holding the substrate at such a low temperature that the initial surface mobility of the vapour atoms was inhibited and hence the required non-crystalline form obtained.

The wide range of uses of thin films is also clearly shown. The electronics industry is an obvious candidate in these days of planar technology but the dielectric properties of oxides and the low temperature coefficient of electrical resistance of both alloys and mixed metal/dielectric structures have been

made use of as capacitors and resistors for many years. A new area of growth is that of integrated optic circuits in which optical signals are processed as they pass along thin surface films of dielectrics, equivalent to flat optical fibres. This is, in some ways, an extension of the already long established use of thin films as blooming coatings on optical lenses and in multi-layer interference filters. Here too there have been many recent advances and new problems uncovered particularly by the use of high energy density optical beams from laser sources. Twenty-five years ago the coating of the Mount Palomar reflector with aluminium was a major task, how many mirrors, prisms and lenses have received a thin film coating since then?

In many areas the future prospect for thin films looks good. Taking only two examples the journal reports significant advances in the thin-film-transistor and in film based solar power cells. By itself the t-f-t cannot compete with silicon technology but for a flat plate television display there is a need for a simple, cheap, mass producible, amplifying device at the end of each (x,y) coordinate. Here thin film technology can play its part and the problems that have been encountered in developing a reliable device are being overcome. The present high cost of solar cells is primarily due to the single crystal silicon used as the starting material. In order to bring the cost down to be comparable with that of conventional power supplies by 1986, as has been proposed, non-single crystal material will have to be used, and in the form of thin layers. The journal reports that thin film technologies should achieve the target on time.

It will be interesting to see what is reported in the 100th volume of *Thin Solid Films*. Perhaps by that time scientists will not be so reticent about the future. For the present the Editors are to be congratulated in giving their authors the chance to look round the field and giving us an end of term report. □

Robert M. Hill is Reader in Physics at Chelsea College, University of London.

Antibodies to probe development

from John K. Heath

THE favoured method in recent years for investigating the cell surface of the early mammalian embryo has been the use of antisera raised against specific embryonic tissue, or teratocarcinoma cell lines.

The best characterised of these is probably the antiserum raised against the F9 teratocarcinoma cell line (reviewed by Jacob *Immun. Rev.* **33**, 3; 1977). This antiserum reacts with a fairly restricted set of tissues; the preimplantation embryo, the germ line and other teratocarcinoma cell lines. The fact that a particular antiserum reacts with a particular set of tissues is not in itself very informative about the role of the cell surface in the development of the early embryo. These reagents do however provide basic information on a population of cells at the molecular level, and often show that cells which are morphologically homogeneous are unexpectedly heterogeneous in their expression of cell surface antigens. F9 antiserum, for example, reacts with nearly all F9 embryonal carcinoma cells (EC cells) but with many fewer, morphologically identical PCC4 EC cells (Seisner *et al. Devl Biol.* **61**, 20; 1977). This and other observations are potentially important because 'classical' experimental embryology is usually based on histological analysis, and has generally classified cells into morphologically discrete types. This has led to a tendency to view the cells in a developing organism as moving through rigidly defined discrete stages. The use of antibodies not only allows one to extract more information on the effects of classical embryological manipulations, but poses questions which would never have arisen if cell morphology alone were used as a guide.

Conventional antisera can only be raised against those antigenic determinants which elicit sufficient response to produce detectable levels of antibody in the serum. Many potentially interesting molecules therefore escape detection. Conventional antisera are also often a complex mixture of antibodies directed against different determinants.

The recent development of 'hybridomas'—cell lines producing specific monoclonal antibodies—provides a way round these difficulties. If spleen cells from an immunised animal are fused with a myeloma cell line some of the resulting hybrid clones secrete antibody directed against the original immunogen (Kohler & Milstein *Nature* **256**,

295; 1975). Each individual clone is descended from a single antibody-producing cell in the original animal and so makes antibody directed against a single antigenic determinant. These monoclonal antibodies promise to revolutionise the study of the early mammalian embryo, amongst many other systems. They dispense with many of the technical drawbacks of conventional antisera, and allow one to find molecules hitherto undetectable on the cell surface. The first papers applying monoclonal antibodies to the study of the early mammalian embryo are now available (Stern *et al. Cell* **14**, 775; 1978; Willison & Stern *Cell* **14**, 785; 1978) and many more will doubtless follow. The results of these studies again show up unexpected differences in the expression of cell surface molecules, which might not have been predicted by classical experimentation. In making hybrids between spleen cells and myelomas one is effectively sampling the whole antibody response to a complex immunogen, and it is therefore possible to make antisera directed against very minor determinants in a population of cells. Indeed, the original immunogen used by Stern *et al.* was a preparation of lymphocytes, partially enriched for T cells. One of the clones resulting from the fusion made a product which reacted with a minor subpopulation of lymphocytes (about 2%) but reacted with EC cells to a much greater extent. This illustrates the value of screening the products of hybrids against a panel of cell types other than the original immunogen. Further investigation showed that the antibody produced by this particular clone reacted with sheep red blood cells, brain and kidney as well as with some somatic component of the testes. The species and tissue distribution suggested that this antigen was similar to Forssman antigen, and the case for this was strengthened by the observation that the antigenic determinant was carried on a glycolipid molecule.

When attention was turned to the preimplantation embryo however, the distribution of this single determinant turned out to be most unusual. The determinant first appeared on the late morula stage (16–32 cells), but could not be found on all embryos. The amount of staining increased as the trophectoderm formed, but again not all trophectodermal cells reacted with the antiserum, and there was no obvious pattern of reactivity common to all embryos. In some cases it appeared as though only one or two cells in the trophectoderm carried the determinant. If the embryo was allowed to grow out

in culture (mimicking the process of implantation) the determinant disappeared from the trophectoderm completely.

The strange distribution of this determinant does not fit in with any preconceived notions of trophectodermal formation or differentiation. It is possible that the expression of this determinant is related to the cell cycle, or that the molecule is perhaps somehow masked in the cell membrane. In any case these findings call for a number of experiments which would not even have been conceived or made possible without antibodies.

Many embryologists believe that the cell surface has an important role in development. There is, unfortunately, very little direct evidence to comfort them. Perhaps monoclonal antibody technology will not only stimulate experiments to reassure the faithful, but help to challenge a number of long held beliefs about developmental processes in general. □

History of biology at the Naples Zoological Station

from Robert Olby

FEW biologists can be unaware of the important contribution which the Stazione Zoologica di Napoli has made over the years since the first visiting scientist came to work there in 1873. As Theodor Boveri declared—"How many prosperous investigations, how much joy of discovery, has this house known!" Established by the wealthy German zoologist and student of the origin of the vertebrates, Anton Dohrn, the station was intended to foster internationally the study of marine organisms. As an enthusiast for the Darwinian theory, Dohrn believed that comparative studies of the development of higher and lower marine organisms would provide the key to the phylogeny of the vertebrates. In fact the Station has been the site for researches of very wide significance in a variety of fields. They include Boveri's classic researches into nucleo-cytoplasmic relations and the individuality of the chromosomes in the sea urchin egg, and Otto Warburg's measurements of the respiration rate of eggs before and after fertilisation, a landmark in the introduction of physiology into development biology. Others who carried out researches at the Station were

John K. Heath is in the Department of Zoology, University of Oxford.

* Held at the Naples Zoological Station Annex on the island of Ischia from July 1–15.

August Weismann, von Uexküll, T. H. Morgan, E. B. Wilson, Albert Szent Györgyi, Jean Brachet, Marcel Florkin and J. Z. Young.

Anton Dohrn had financed the Station from four sources—receipts from visitors to the Aquarium (then as now housed in the Station) contributions from institutes and governments who rented 'working places' for visiting members from their institutes, proceeds from the sale of preserved marine animals, and contributions from patrons. This private status of the Station came to an end as a result of reorganisation under a new Administrative Council formed in 1975. Two directors have served under this Council, Alessandro Barlaam and the present director, Alberto Monroy.

Both directors were aware of the importance of the archival material at the Station. Several years ago the archivist Christiane Groeben sorted and catalogued the scientific correspondence, photographs, and other documents with the aid of funds from the Dohrn family Foundation and the Deutsche Forschungsgemeinschaft. In 1974 an exhibition of this material was held at the Station; copies of the excellent catalogue are still available.

As the existence of this remarkable collection of some 15,000 letters is becoming more generally known Groeben's task is growing. The historical work of the Station is being expanded with the formation of a Research Group consisting of Bernardino Fantini of the Institute of Genetics in Rome, Christiane Groeben and the Station's librarian Walter Groeben.

The aim of the group is to promote the history of embryology, cell biology, biochemistry, neurophysiology and taxonomy, with special attention to the social and cultural role of research institutes. In addition to the organisation of the scientific correspondence at the Station, scientists who have had connections with the Station will be interviewed. It is planned to start with J. Brachet, A. Monroy, G. Montalenti, M. Aloisi and J. Z. Young.

At the first of a proposed series of biennial summer schools, this year on biochemical and molecular embryology*, Jean Brachet gave a spirited account of his involvement in chemical embryology. The ways in which the subject of the distribution of nucleic acids in the cell was affected by the lack of congruence between the disciplines of biochemistry and histochemistry, and the impact this had on his scientific career came over clearly.

Robert Olby is in the Division of the History and Philosophy of Science at the University of Leeds.

Petrified cotton?

from F. L. Vogel and J. N. Zemel

A NOVEL material has just made its debut which provides an artificial counterpart to natural petrified materials which preserve the fine detail of the original perishable structure. Natural petrified wood for example, is formed when the original wood is infiltrated by silicious matter which converts it into a stony substance preserving the original structure. There are no comparable processes used industrially that convert organic materials to a more permanent form. The recent paper by T. L. Ward and R. R. Benerito (*Thin Solid Films* 53, 73; 1978) is therefore particularly interesting. They describe the unusual transformation of ordinary cotton fibre into a glassy material reminiscent of the permineralisation of wood.

In the Ward-Benerito process, cotton fabric is first soaked in a sodium plumbite solution after which it is dried, placed on a glassy substrate and heated to temperatures as high as 600 °C. During heating, the treated cotton initially carbonises to a black colour. As the temperature is raised, it picks up silicious material from the substrate and becomes yellow. At 600 °C the carbon skeleton is impregnated with silica and is converted to a glassy substance reproducing details of the original fabric. The authors demonstrated the results of this process with many artefacts including labels and decorations on bottles, adhesives between glasses and free standing thin films.

These results are interesting as they indicate the existence of a novel material. Still more interesting perhaps is the origin of the process which produces it. Ward and Benerito examined the composition of the plumbite-treated material using a variety of analytic techniques such as X-ray diffraction, optical and electron microscopy (TEM and SEM), ESCA, EDAX, DTA, X-ray fluorescence and atomic absorption. These studies indicated that lead was uniformly distributed in both the unheated and the transformed cotton. Furthermore, it was observed that the residual carbon concentration in the charred cotton

was six carbon atoms to one lead atom. Since lead is not known to form a carbide, this result may be taken as an indication of an interstitial or intercalation compound described in a recent popular review (Fischer & Thompson *Phys. Today* July, 1978). This conjecture is supported by the following observations.

When cellulosic materials are charred, a high defected hexagonal layered structure forms. While this is not a true graphite, it may be sufficiently similar that some degree of intercalation is possible.

The process of interstitial insertion can in itself improve the crystal perfection of the hexagonal layered structure resulting in a closer approach to a graphitic material.

The relatively low temperature of the process suggests that the diffusion occurs by an interstitial mechanism.

Silicates are known to form layered compounds, one common example of which is mica.

The mechanism suggested by this combination of facts is that intercalation of lead separates the spacing between the roughly formed hexagonal layers sufficiently to allow interstitial diffusion of the silicious materials from the substrate.

Neither the 'petrified cotton' nor the process to produce it arose in response to a stated need. As such, it is a fascinating example of a fundamental study of a technical substance. Since long term stability of cellulosic materials such as wood, paper and cotton is notoriously poor, it might be expected that permineralisation of these substances by the Ward-Benerito technique could be used. Diverse examples of this would include document preservation, converted wood as a decorative building material, or lacy *objets d'art* that could be produced no other way. As with any new material, uses will probably be found that cannot be foreseen at this early stage.

F. L. Vogel and J. N. Zemel are in the Department of Electrical Engineering and Science, University of Pennsylvania.

His modest description of his work as the correction of a series of errors was accompanied by his strong claim for the importance of his work on the transmission of information from DNA to the site of protein synthesis by long-lived RNA molecules. His statements on this subject are to be found

in his Weizmann Memorial Lecture, *The Biological Role of the Nucleic Acids*, delivered in April 1959 and published in 1960. The discovery of soluble RNA and microsomal RNA did not solve what "for the geneticist and immunologist [was] the major problem", wrote Brachet, "the mech-

anism of specific protein synthesis." Such a mechanism could be visualised if it were assumed that the "nucleus produced an intermediary substance involved in protein synthesis; there is little doubt that this intermediary is RNA...". In contrast to the *Escherichia coli* model with a short-lived messenger, Brachet's model, based on the alga *Acetabularia*, necessarily assumed a long-lived messenger, because of the long duration of protein synthesis in enucleated fragments, and this model represents more closely than the *E. coli* model, the situation in higher organisms.

Reading Brachet's Weizmann Lecture today one is struck by the scattered distribution of specific statements on the nature of 'template' (that is messenger) RNA. This may in part account for the lack of appreciation of these *Acetabularia* studies in the history of molecular biology.

The relation between chemical structure and morphological organisation was explored in several papers, most notable of which was the one by Jacques Roger of the University of Paris. He outlined the early history of the concept of an 'organic substance' as a state of chemical combination responsible for vital phenomena independent of any morphological organisation in the anatomist's and microscopist's sense of the word. The concept of a glutinous, mucilaginous, albuminoid, and later a colloidal substance can, Roger claimed, be traced back to this organic substance. Equally, the claim that the chemist can study this substance to the advantage of our understanding of vital phenomena can be traced back to eighteenth century writings on the organic substance—those of Bourguet, Buffon, Rouelle, Venel and Lamarck. In other papers the importance of the recognition of definite limits to the miniaturisation of morphological structures was stressed. Maxwell in 1875 was the first to assert that the dimensions of protein molecules set a lower limit to the postulation of such organised particles as the gemmules introduced by Charles Darwin in his hypothesis of pangenesis. Fantini emphasised the novelty of the concept of three-dimensional, specific structure in macromolecules determined by sequence, and he contrasted this with the vaguer conception of shape to be found in colloid science.

The establishment of this Research Group in Italy is an important event in the history and philosophy of science. Hitherto work in the subject has tended to be dominated by Anglo-American activities. The programme initiated by the Research Group should not only act as a catalyst to the growth of the subject in Italy, but should also stimulate such work in the area of the

life sciences, as befits a group associated so closely with the international activities of the Zoological Station. □

Globular clusters

from Virginia Trimble

THE globular clusters comprise the oldest stellar population readily available to us for study. They ought, therefore, to be able to provide the answers to various questions in galactic and stellar formation and evolution—if only we knew which questions to ask. The chief difficulty lies in deciding whether they are the rare (about 10^{-4} of the observed mass in the Universe) survivors of a population of once-common, fundamental building-blocks from which larger units (galaxies and clusters) have arisen, or the products of early fragmentation in previously-existing gas clouds that we might call protogalaxies.

Uncertainty on this basic point underlay and sometimes befuddled discussions of new observations and calculations at a recent NATO Advanced Study Institute on Globular Clusters.* For instance, the average heavy-element abundance of a population of globular clusters increases with the mass of the galaxy to which it is attached (S. van den Bergh), and, within at least the Milky Way and the Andromeda Galaxy, abundances range systematically from nearly solar at the centre, through a steep decline 5-10 kpc out, to one-twentieth of solar or less in the halo (with large scatters at all positions; L. Searle). But is this trying to tell us something about how star-formation rates (and, therefore heavy element production) depends on gas density or something about how dynamical friction juggles units of varying mass and degree of central concentration?

Dynamics of stars within individual clusters can be considered almost without regard to their origins, because at least the inner parts are relaxed. The identification of seven X-ray sources (including five bursters; W. Lewin, G. Jernigan) with globular clusters lends extra urgency to calculations of core evolution and collapse (L. Spitzer), although recent estimates of the surface area required to produce the bursts and the relative timing of optical and X-ray emission in one burst now strongly suggest that the source is an accreting neutron star rather than a massive black hole formed by core collapse. Core evolution may still have

to be invoked to put a neutron star into a suitable X-ray-producing binary system, as it becomes increasingly clear that the clusters are deficient (relative to stars in the galactic disk, at least) in binary systems over a considerable range of parameters (J. Gunn, R. Griffin, M. Liller, E. Budding). But all is not lost, even if core collapse never occurs (J. Katz), as it may be possible to get the observed X-ray flux from single neutron stars occasionally passing through the stellar winds of red giant stars (A. Fabian). The amount of ambient gas implied by this is well below the $0.1\text{--}1.0\text{ M}_{\odot}$ seen or set as upper limits from HI and HII studies of several clusters (G. Knapp, W. Liller). Some of these limits are, however, uncomfortably low for standard theories of mass loss from evolving stars and gas escape from clusters (D. Faulkner).

Progress has been made on several traditional, vexing, rather technical issues. The wide giant branch in ω Cen is undoubtedly to be attributed to real abundance variations, and because CN line strength is generally correlated with Ca, Fe, and Sr abundances (J. Norris), the variations must have been intrinsic to the gas from which the stars formed, and not just produced by mixing within individual giants. Real variations in CNO abundances occurring among and within other clusters (especially 47 Tuc; A. Rodgers), are not always uniquely correlated with variations in Fe and heavier elements (J. Frogel, E. Persson), and may also be primordial (R. Kraft); but they are not the whole answer to the traditional 'second parameter' problem.

The second parameter(s) is whatever is responsible for the lack of the theoretically-expected correlation of the appearance of a cluster's HR diagram with its heavy element abundance. The most conspicuous effect is the occurrence of both red and blue horizontal branches in metal-poor clusters, especially in the outer parts of the galaxy (R. Zinn), when only blue ones would be expected. High CNO (either primordial or mixed up from the core) would cure this, but CNO cannot deal (V. Castellani) with a more subtle effect seen in some far-halo clusters (R. Schommer) in which the red-giant branch is too red for the metal abundance implied by the main-sequence turn-off point and integrated cluster spectra. HR diagram morphology is also affected by helium abundance, cluster age, mass loss from red giants, and stellar rotation. It will probably be necessary to include the contributions of at least two 'second parameters' to match the full range of observed HR diagrams. Rotation is apparently also important (A. Sweigart) in finally making it possible to under-

*Held on 7-18 August at the Institute of Astronomy, Cambridge.

stand how processed material can be mixed up from the hydrogen-burning region to the surface of low mass stars by means of meridional circulation.

No consensus now seems possible on the meaning (or, in some cases, even the reality) of colour and abundance gradients within clusters other than ω Cen (K. Freeman; M. Chun; G. Da Costa). The question of whether or not the absolute brightnesses of RR Lyrae variables are correlated with their metal abundances is 'controversial.' It is important because these stars are one of our standard candles, whose brightness partially determines the extragalactic distance scale, and so Hubble's constant and the age of the Universe. W. Liller

and V. Clube strongly advocated a relationship (partially based on recent photometry by B. Carney) in which the RR Lyraes in clusters of lowest metal abundance are about one magnitude fainter than those in clusters of high Z. A constant brightness for these stars, independent of cluster composition, was equally strongly supported on both observational and theoretical grounds by P. Demarque, R. Kraft, and F. Hartwick.

The various problems of globular cluster astronomy are probably in as good shape as we can expect, considering that (although some of the clusters were recorded by Messier) the first classification of them by radius and central concentration was begun only about 50 years ago by H. Sawyer Hogg, who was an active participant at the Institute. $\square$

Virginia Trimble is at the Astronomy Program, University of Maryland.

Bacterial gene into monkey cells

from John Jenkins

ONE of the long term aims of genetic engineering recently came a step closer to reality, with the stable propagation in mammalian cells of a hybrid DNA molecule containing a mammalian tumour virus chromosome and a bacterial gene. The successful development of such vectors for introducing and propagating genes in mammalian cells offers the hope of eventually being able to study the control of expression of eukaryotic genes in a system not too far removed from their natural environment. Even further in the future is the possibility of using such vectors to introduce particular genes to correct inborn genetic defects.

The idea itself is not new. In 1972, Jackson *et al.* (*Proc. natn. Acad. Sci. U.S.A.* **73**, 2904; 1972) described the construction of a hybrid molecule from simian virus 40 (SV40) and bacteriophage λ DNA, and such hybrids can replicate in suitable host cells (Ganem *et al.* *Cell* **7**, 349; 1976). The problem is that replication of the hybrid virus kills the host. A new development gets round this problem by tailoring or 'crippling' the SV40 so that the genes responsible for cell death are cut out before the hybrid is made.

SV40 can behave in two different ways when it infects a cell, and its genome can be divided into two regions on this basis. In rat cells, the entire virus genome is inserted into the host

chromosome and the cells change many of their growth characteristics. This transformation requires the expression of only about half the SV40 genome making up what is known as the early region. The second type of infection, of monkey cells for instance, requires expression of most of the genome. In this case, the remainder of the viral genes are functionally expressed and about 36 h after infection mature virus particles appear, followed by the lysis of the host cells. Cell types that support this lytic cycle of infection are termed permissive, while those which transform but fail to produce mature virus are called nonpermissive.

In order to cut out the genes responsible for lytic infection, (Hamer *et al.* *J. molec. Biol.* **112**, 155; 1977) cleaved the circular virus DNA at two points within the late region, using the restriction enzymes *HpaII* and *EcoRI*. The large fragment from this, containing the origin of DNA replication and the early region genes, was ligated to a DNA fragment containing an *E. coli* tRNA suppressor gene. In order to get large amounts of the hybrid DNA it was necessary to first grow it up in monkey cells in the presence of a helper virus which possessed functional late region genes. Purified hybrid DNA from this lytic cycle was then used in the present series of experiments (Upcroft *et al.* *Proc. natn. Acad. Sci. U.S.A.* **75**, 2117; 1978).

The hybrid DNA was found to transform nonpermissive rat cells, and analysis of total cellular DNA from

infected cultures showed the presence of about three copies of non-integrated hybrid genome per cell. Some copies might have integrated with the host DNA and been maintained as such but the evidence for this was not conclusive. Infected monkey cells, which with wild type SV40 support the production of mature virus, were found to contain amounts of nonintegrated, superhelical hybrid DNA which was stably propagated when the cells were cloned.

There are several advantages to this system of nonlytic propagation using a crippled vector. First, the recombinant DNA is never expressed as an infectious virus particle simply because it does not possess the necessary genes. This satisfies some of the safety considerations involved in this type of work. Second, since assembly of a complete particle is not needed, the amount of DNA that can be 'carried' on the SV40 vector is probably increased, and third, the host cells remain viable, which offers the chance of long term studies on the maintenance and expression of 'foreign' DNA sequences in a mammalian system.

Hamer *et al.* looked for expression of the bacterial gene in the original lytic infection and found that although the sequence was transcribed into RNA, no functional suppressor tRNA was detectable. This is not surprising in view of the different types of processing of tRNA that probably take place in bacterial and eukaryotic cells (Knapp *et al.* *Cell* **14**, 221; 1978; O'Farrell *et al.* *Nature*, **274**, 438; 1978). This may not be a problem when the gene 'insert' is from a more closely related donor. Rumour has it that correct transcription and expression of globin genes cloned in an SV40 vector in mammalian cells has been achieved in Paul Berg's laboratory at Stanford.

There are a variety of sequences that could usefully be worked on in a system of this kind, but many of them will probably have to wait until more has been done to reduce the possible hazards of working with a tumour virus. The vector in these experiments could still transform cells, and it may be necessary to remove this ability before the SV40 is sufficiently 'crippled' to use with genes that might confer a selective advantage on the hybrids.

Other routes exist for getting genes into mammalian cells, for example, by cell fusion and selection (Szybalski *et al.* *Natn. Cancer Inst. Monogr.* **7**, 75; 1962; Littlefield *Science*, **145**, 709; 1964) and such genes may be expressed in adult animals (Illmensee *et al.* *Proc. natn. Acad. Sci. U.S.A.* **75**, 1914; 1978). But in terms of simplicity and flexibility the tumour virus vector could be a winner. $\square$

John Jenkins is working in the Nucleic Acid Laboratory, Imperial Cancer Research Fund, London.

Forty years of plant mitochondriology

from A. L. Moore

It has taken approximately 40 years since the first paper was published describing their oxidative capacities for a symposium to be held specifically devoted to plant mitochondria.*

Although the cytochromes of plant mitochondria are spectrally and functionally similar to those of animal mitochondria, there is still confusion about the role of the multiple *b* cytochromes in electron transport. B. T. Storey (University of Pennsylvania, Philadelphia) suggested that cytochromes *b*-556 and *b*-560 are part of the respiratory chain whereas cytochrome *b*-565 only associated with the main chain in the energised state and in the presence of antimycin A. This suggestion was based upon the kinetic behaviour of *b*-565, in particular its complex oxidation/reduction kinetics during coupled substrate oxidation. However a few papers were presented which disagreed with this interpretation and it was suggested that a possible Q-cycle mechanism could explain the complicated nature of *b*-565 reduction under oxidative conditions.

P. R. Rich (University of Cambridge) gave a detailed analysis of the paramagnetic centres associated with higher plant and *poky N. crassa* mitochondria. Of particular importance was the finding that no EPR-detectable component was directly attributable to the alternative oxidase and furthermore it was suggested that there is no evidence for a direct involvement of iron sulphur centres or transition metals in a paramagnetic state in the terminal oxidation system. Although the HiPIP and ferredoxin-type centres are analogous to their mammalian counterparts, an EPR signal characteristic of a high-spin ferric iron with a *g* value at 4.3 was independently reported by Rich and M.-F. Henry (University of Amsterdam). Henry presented evidence in favour of this iron-containing protein's being involved in the cyanide insensitive respiratory pathway although the redox behaviour of this signal under conditions in which the alternative oxidase was operating was questioned. Obviously further work on the function of this component is required before its possible role in alternative oxidations can be better understood.

The topic which generated much interest and led to lively round-table discussions was the relationship between the malate oxidoreductases and NADH dehydrogenases. The pro-

cess of malate oxidation is much more complex in plant than in animal mitochondria. Malate can be oxidised by two malate oxidoreductases and the NADH produced may be oxidised by at least four different NADH dehydrogenase systems. One of these dehydrogenases is located on the outer surface of the mitochondrial inner membrane.

Although there is general agreement that the oxidation of malate is mediated by an NAD⁺-linked malic enzyme and malate dehydrogenase, the exact location of these enzymes is still controversial. J. M. Palmer (University of London) took the view that the oxidation of malate can provide NADH for oxidation by both the internal and external NADH dehydrogenases and that malate reduces external NAD⁺ primarily by the activity of externally located malic enzyme and malate dehydrogenase. In contrast, however, J. T. Wiskich (University of Adelaide) supported the view that the external NADH dehydrogenase was not involved in the oxidation of malate in the absence of added NAD whereas in its presence a transmembrane transhydrogenase has been postulated to operate, which is considered to transfer reducing equivalents from internal NADH to external NAD thus connecting internal malate oxidation to the external dehydrogenase. D. A. Day (University of California) presented evidence on the malic enzyme distribution in potato mitochondria from which he concluded that malic enzyme is not present in the intermembrane space of these mitochondria and furthermore that all malate oxidation occurs in the matrix. Perhaps in this respect, the most interesting findings reported at this symposium were the detailed investigations of M. Neuburger (Université de Grenoble) on the transport of NAD⁺ through the inner membrane of purified mitochondria. Convincing evidence was presented which indicated that highly purified intact plant mitochondria possess a NAD-translocator. Most workers readily agreed that the existence of such a carrier would explain many of the conflicting reports in the literature on malate oxidation and the further characterisation of this carrier is eagerly awaited.

The sessions on the molecular and cellular aspects of cyanide resistant respiration attracted considerable attention. The exact nature of the alternative oxidase has remained elusive for many years. However, as W. D. Bonner (University of Pennsylvania, Philadelphia) pointed out, a considerable amount of evidence has accumulated which allows a fuller understanding of its branchpoint from the main respiratory pathway and the nature of the products of the alternative oxidase with

oxygen. Bonner suggested that a modified protonmotive ubiquinone cycle satisfies the kinetics of cytochrome interactions, and that the alternative pathway accepts electrons non-protonmotively from this cycle. This proposal suggests that a pool of ubiquinone in the succinic dehydrogenase region of the respiratory chain is performing as the branchpoint for the alternative pathway. Bonner concluded that the alternative oxidase could either be a species of ubiquinone or an EPR-silent metalloprotein that facilitates the reaction of the ubiquinone species with oxygen. In this regard a great advance in our knowledge of the nature of the alternative oxidase was obtained from the studies of Rich and S. Huq (Imperial College, London) who presented results on the isolation and partial purification of the alternative oxidase from *Arum maculatum*. Both workers used a hydroxamic acid-sensitive quinol oxidation as an assay for the oxidase. Remarkably, Rich found that in a deoxycholate solubilised preparation the quinol oxidase activity remained very stable, which seems to suggest that the lability of the oxidase in intact mitochondria may be due to the connection between the dehydrogenases and oxidase, and not the alternative oxidase itself. Both preparations contained minimal amounts of cytochromes *aa₃*, *c* and *b*. It was of particular interest to note that the overall product of oxygen reduction by the quinols was water.

A. M. Lambowitz (University of St. Louis) reviewed the current status of the biosynthesis of the alternative oxidase in *Neurospora* mitochondria. An important development in this area has been the isolation of antimycin-sensitive mutants with defects in the biosynthesis of the alternative oxidase. All cytochrome system activities seem unaffected in the mutants and the only apparent defect is in the induction of the alternative oxidase. Lambowitz suggested that the results indicate that there are at least two polypeptides unique to the alternative oxidation system. □

A. L. Moore is a Rank Prize Research Fellow in the Department of Plant Sciences, King's College, University of London.

Erratum

In the article 'Halogenated hydrocarbon effects' (*News & Views* 274, 533; 1978), in the last paragraph, column 1, on page 534, the first line should read "Both the PCBs and the polybrominated biphenyls (PBBs) induce hepatic effects". In the second paragraph, column 2, line 4 ff. should read "Poland . . . believes this to be a receptor for TCDD and consequently for other inducers of aryl hydrocarbon hydroxylase . . .".

*The First International Symposium on Plant Mitochondria was held at the Luminy campus in Marseilles on 31 July-4 August.

review article

ASSESS II: a simulated mission of Spacelab

H. M. Wegmann¹, R. Hermann¹, C. M. Wingett², M. de Muizon³, D. Rouan³, P. Léna³, J. Wijnbergen⁴, H. Olthof⁵, K. W. Michel⁵, Ch. Werner⁶, F. Melchiorri⁷, B. Melchiorri⁷, V. Natale⁷, R. Falciani⁸, L. R. Smaldone⁸, E. Bussoletti⁹, J. Crawford¹⁰, P. Rothwell¹⁰, M. Taylor¹⁰, J. E. Beckman¹¹, D. Dale¹¹, J. Schmitt¹¹ & J. de Waard¹²

ASSESS II, a joint exercise of NASA and ESA, was set up to enable scientists to become accustomed to an atmosphere-free, gravity-free working environment. This simulation of Spacelab conditions has provided new data and a valuable insight into the possibilities and limitations of such a mission.

IN THE 1980s many scientists in Europe and the USA will be able to fly experiments in space using the Spacelab now under development by the European Space Agency (ESA). The experiments will cover a wide range of disciplines and will often fly on a single mission together with equipment for engineering and industrial applications. It is important for would-be experimenters, and ESA, to learn early those operational techniques specific to Spacelab which will enable scientists to gain maximum benefit from its advantages of an atmosphere-free and gravity-free environment.

In 1975 NASA and ESA conducted a joint exercise, ASSESS I, using a Convair 990 research aircraft, the Galileo II, belonging to NASA's Ames Research Center, in a preliminary simulation of a Spacelab flight. Four scientists were confined to the aircraft for a week, during which six research flights were conducted, with a multi-experiment payload (three USA and three European), aimed at astronomy and stratospheric studies. The results of ASSESS I were encouraging, in that they showed how experimentalists with only a few weeks' specialised training could act as proxy operators (payload specialists) for a range of experiments not necessarily from their own fields of research. However, the project was assembled rather hastily, and was not as realistic a simulation as desired. Experiments were not organised ergonomically into a payload, making the work of the payload specialists exceptionally difficult. No attempt was made to use a common integration site, with specially prepared ground support equipment, and the ADDAS onboard computer system for handling scientific data could not be used satisfactorily because of lack of time to prepare and check software.

ASSESS II, flown in mid 1977, aimed to improve the realism of the simulation, and provide another chance for scientists to acquire new data. The same aircraft was used, with six American and six European experiments on board, run by four payload specialists. The mission was longer, with nine research

flights each of 6-h duration carried out in a 10-d period. Payload specialists were confined within the aircraft, or in a juxtaposed sleeping trailer, communicating to the outside world only by telephone or radio. Here we describe how the mission was organised, and present some of the initial data from the European experiments.

Choice of experiments

The ESA selection committee based its choice of experiments to fly on (1) its scientific value; (2) the state of experimental readiness; there was just over a year between experiment selection and flight, so it was preferable to use an experiment which did not have to be built up from scratch; (3) its ease of operation: using payload specialists to run three experiments each, some highly complex, meant avoiding apparatus with many fine tuning adjustments, switches to throw continually, and widely separated pieces; and (4) the similarity to the types of experiments envisaged for Spacelab.

The full complement was chosen for mutual compatibility to avoid, as far as possible, clashes in the mission timing and trajectory plans. The selected experiments were: A threefold infrared astronomy experiment, based on a 32-cm infrared telescope (Meudon). This carried a continuum photometer for observing interstellar dark clouds (Groningen), and a line spectrometer for interstellar sulphur in HII regions (Garching); a television camera to study infrared airglow from noctilucent terrestrial clouds (Southampton); a sub-millimetre spectrophotometer to measure the Sun's chromospheric brightness temperature (Naples-Florence); a LIDAR system to measure aerosol distribution in the troposphere (Oberpfaffenhofen); medical probes to study diurnal and stress-induced behaviour of the payload specialists (Bad-Godesberg); and test equipment to study the electromagnetic environment, and help remedy earthing problems (Noordwijk). (This was operated by the crew and used also on behalf of the USA payload.)

Italian, French, Dutch, German and British groups were represented as for ESA satellite programmes, and as intended on Spacelab.

Experiment preparation

Build up of the ASSESS II experimentation was begun at the experimenters' home institutes in late 1976; the payload specialists received two to three weeks' initial training on each

1. DFVLR Institut für Flugmedizin, 53 Bonn, FRG.

2. NASA Ames Research Center, Moffett Field, California 94035.

3. Groupe Infrarouge Spatial, Observatoire de Meudon, France.

4. Department of Space Research, University of Groningen, Netherlands.

5. Max-Planck-Institut für extraterrestrische Physik, Garching, FRG.

6. DFVLR Institut für Physik der Atmosphäre, 8031 Oberpfaffenhofen, FRG.

7. Infrared Group IROE-CNR 50145, Firenze, Italy.

8. Capodimonte Astronomical Observatory, 80131 Naples, Italy.

9. Institute of Physics, University of Lecce, 73100 Lecce, Italy.

10. Department of Physics, University of Southampton, UK.

11. ESA, ESTEC, Noordwijk, Netherlands.

12. ESA-SPICE, Linder-Höhe, Porz-Wahn, FRG.

experiment in this period. In January 1977 all the experiments were gathered at the ESA/SPICE facility at the DFVLR, Porz-Wahn near Köln. Here they were integrated electrically and mechanically into a mock-up of the aircraft, and simulation tests of the European half of the payload were carried out. It was convenient for the principal investigator teams to attend for long periods, but the payload specialists who were on the spot all the time increasingly handled the equipment. During this period they were able to write their detailed procedures for installing and operating the experiments on the aircraft.

The simulation was very realistic, demanding continuous manning of the experiments and of a control room designed and operated as a replica of that at Ames. Two sets, of four 'flights' each, tested the skills of the payload specialists, and their interaction with the ground teams. After this simulation the experiments were moved to California, where the whole (including USA) payload was put together on the ground at NASA Ames, then installed in the aircraft and flown.

The data flights

In late May 1977 the simulation mission of nine flights was performed, followed by a set of 'principal investigator flights', during which the Spacelab-like conditions were relaxed, and more personnel went on board to assist with each experiment. The nine simulation flights took place on schedule with minimum intervals of less than one day for maximum loading on the payload specialists. Communication during the mission was via a control centre, in which the teams running the flight and payload operations lived. The scientists on the ground could

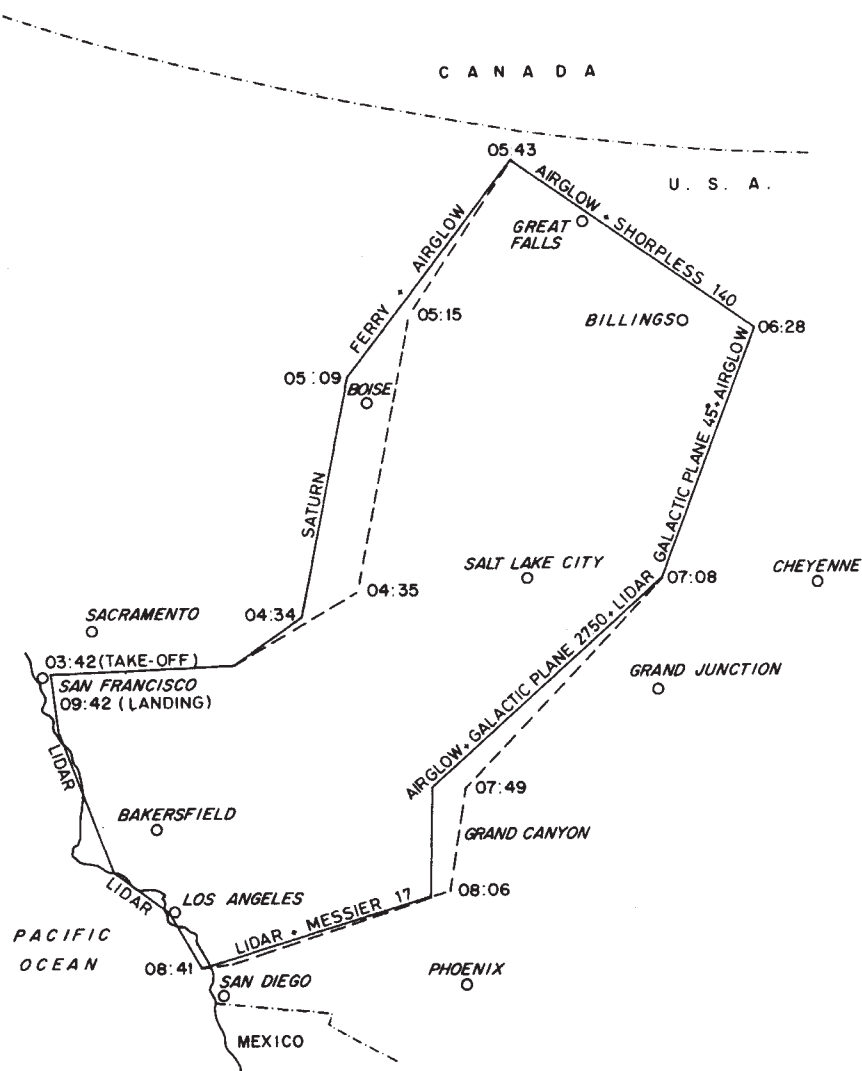
talk to the payload specialists for limited periods, about two hours per day, to advise on experiment preparation in the light of information from the previous data-take. A TV camera enabled some visual information to be passed 'upwards' to the aircraft (only available when on the ground) while a regulated allowance of data was removed from the aircraft between flights, corresponding to a bit rate which can be accomplished by planned telemetry from the shuttle. In practice communication during actual flights was difficult—this was one way in which conditions were significantly worse than those anticipated for Spacelab. Another difference was the successive takeoffs, landings, and refuellings, with their disruptive effects. Figure 1 shows a typical trajectory for a night flight, and gives an idea of the distribution of activities for the European experiments during such a flight. The LIDAR and the solar experiment were kept running for most sections of all the trajectories, providing routine data on the Earth's atmosphere. Two night flights over the Pacific viewed southerly astronomical objects, and the day flights included prolonged manoeuvres over cities and peculiar terrains, for LIDAR and for the USA payload, which was mainly Earth-resource orientated.

The individual experiments, and their results will be presented elsewhere. Here we indicate the organisational complexity of the project, and the payload specialists' tasks.

Infrared astronomy

A 32-cm aperture infrared telescope which had flown on ASSESS I, carried two instruments, a 4-band photometer spanning the wavelength range 31–196 μm , and a Fabry-Pérot

Fig. 1 The trajectory of an ASSESS II night-time flight over the Western USA. Data were taken throughout each 6-h flight, while preparations and post flight checks took between 6 and 10 h longer. Flight tracks for Purple 1, 21 May 1977: solid line, original plan; dashed line, revised plan; all times UT.



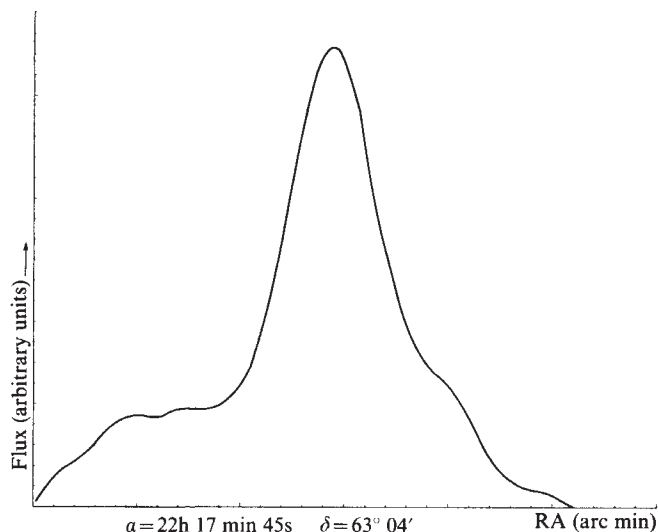


Fig. 2 A scan of the dark interstellar cloud Sharpless 140, obtained with the Groningen photometer in the waveband 115–196 μm (TII filter). The abscissa registers a scan amplitude of 29 arc min in RA using a diaphragm of 4.5 arc min diameter.

line spectrometer, both cryogenically cooled. The payload specialist could switch between them using a flip-flop mirror in the telescope. A sensitive TV camera facilitated finding and locking onto infrared position, by serial spot-follower and gyro-stabilised control loops. These experiments were run throughout the mission by one payload specialist.

Figure 2 shows a typical output scan; new flux maps of the molecular clouds S140(A), R Cr A, and M17 were built up from such scans, and previous measurements of the galactic plane^{2,3} and of sky noise⁴ were revised with increased precision. Spectra at the 33.6 μm wavelength of SIII emission with the Fabry-Pérot system were, however, inconclusive.

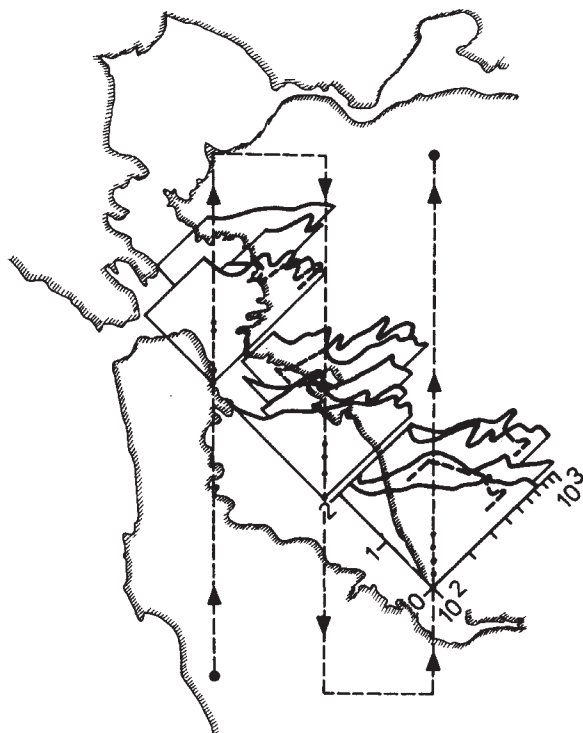


Fig. 3 Mass concentration of aerosol particles (in mg m^{-3}) as a function of altitude over a portion of the San Francisco Bay area, measured by the LIDAR experiment on ASSESS II.

Solar brightness temperature

During daytime flights a Michelson interferometer was fed by a gyro-stabilised coelostat, giving solar spectra in the range 300 μm –2 mm wavelength. The aim was an absolute determination of the chromospheric brightness temperature to within 250 K, for use in physical modelling. The equipment was also operational on all other flights, yielding terrestrial atmospheric emission spectra varying with latitude and altitude, and examining submillimetre sky anisotropy over a 2° field, aimed at the cosmic background radiation⁵.

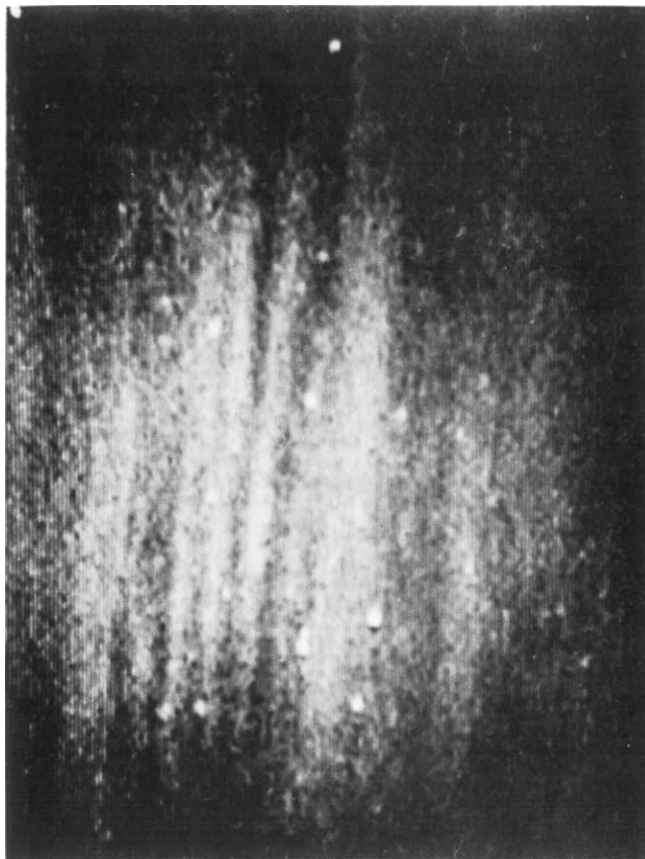


Fig. 4 OH airglow seen in the form of narrow stripes with the TV camera system at 700–900 nm.

Lidar

Figure 3 shows an aerosol distribution diagram for an area south of San Francisco. It shows the effectiveness of the airborne LIDAR, which obtained such data over urban areas (Los Angeles basin and San Francisco Bay) and over very varied terrain (Mojave desert, Rocky Mountains, and the Central Valley of California). The already standard LIDAR technique of measuring the time of flight of back-scattered pulsed near infrared laser radiation emitted from the aircraft, was calibrated using the simultaneous data of Dr P. Russell, taken from the ground at the Stanford Research Institute.

Airglow TV

Mesospheric airglow from the OH layer, ~85 km above the Earth was studied for structure due to winds, atmospheric tides, and gravity waves. The TV technique used was a development of that flown on ASSESS I (ref. 6), recording images in the 700–900 nm waveband every 0.15 s. This compares with 5–10-min photographic exposures from mountain sites^{7–9}. Figure 4 shows narrow-banded mesospheric structures, observed during ASSESS II. During the mission broader bands were commonly seen, also patches, and V-shaped structures. Amid this variety of geometrical patterns, banded structures with tens of km periods and maximum

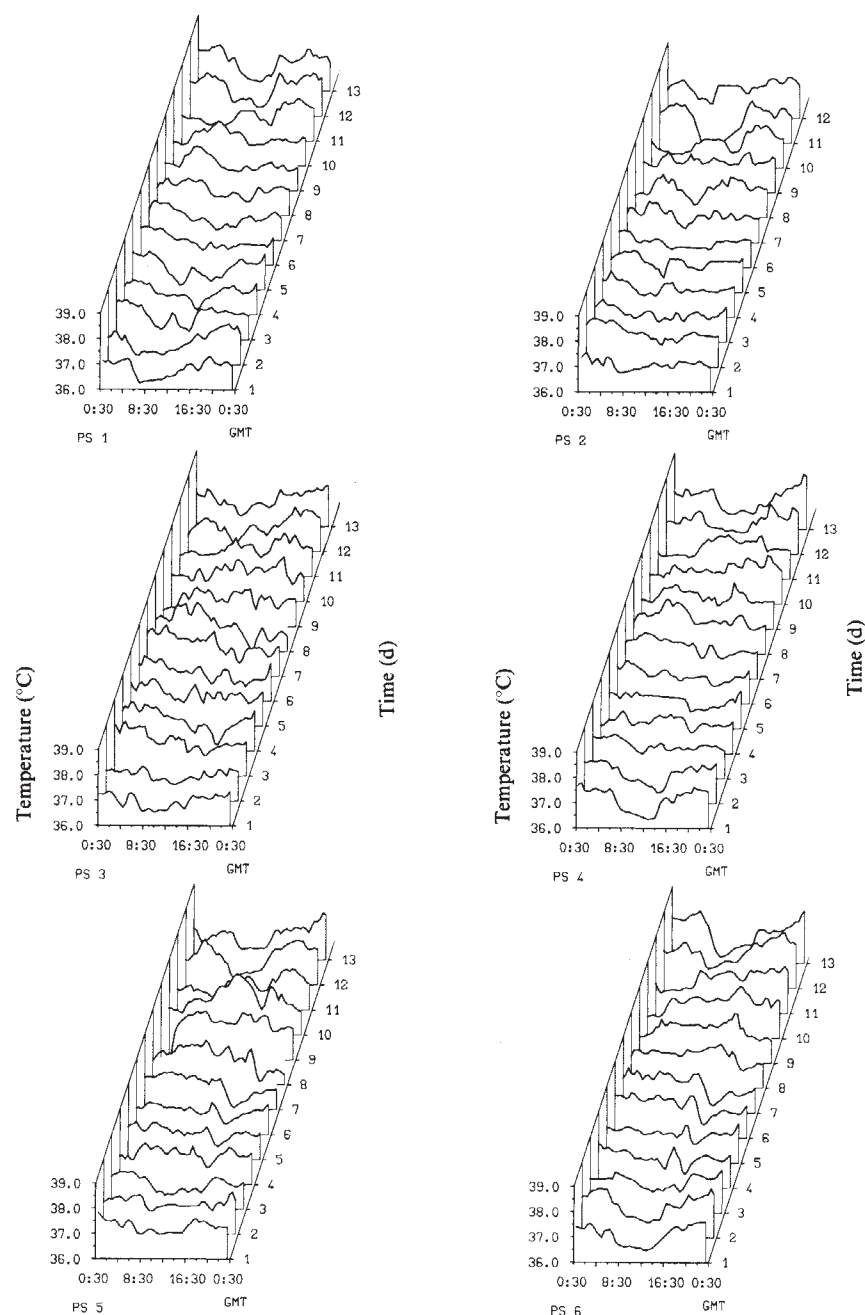


Fig. 5 A survey, abstracted from the medical experiment data, of the temperature rhythms for the six payload specialists (two USA, two European, and two European ground back-up) working on the ASSESS II mission.

extension around 700 km were frequently prominent. The latter seem too small to fit gravity wave predictions¹⁰. The data, a clear improvement on ASSESS I, show that tropospheric storms can cause mesospheric airglow pattern changes, and provide evidence for multiple layers in the airglow, from sets of bands at fixed common angles.

Medical experiment

The medical experiment was aimed directly at Spacelab. Each payload specialist (two Europeans and two Americans flying, plus two European back-ups on the ground) was fitted with a rectal temperature probe, and ECG (electrocardiogram) electrodes, recording continually on to miniature tape-recorders carried on a belt. During sleep the EEG (electroencephalogram) was measured and similarly recorded. Every three hours, except during sleep, a urine sample was collected and stored frozen for subsequent analysis of hormones and electrolytes.

A global look at the temperatures of the payload specialists (PSs) is given in Fig. 5, for two days pre-flight, nine flight days, and two post-flight days. Effects of slipping the working routine

against the 24-h day are seen in the drift of the diurnal maxima and minima, (see PS 1, 5 and 6); the increasingly broken rhythm (see PS 3, also 2 and 4) is due to poor sleep. Full analysis of the 10 parameters recorded will be valuable for planning activities and routines for Spacelab.

Experiment grouping

The infrared astronomy experiments, comprising five experiment racks and the telescope, were run by one payload specialist while the LIDAR, Airglow and Solar experiments were run by the other. Ergonomically this worked quite well, as the latter three were controlled mainly from a centralised panel, and the effort demanded for each group was roughly equivalent. Although considerable progress in reducing trivial switch-throwing had been made since ASSESS I, further automation would have undoubtedly reduced stress on the payload specialists.

Data management

The airglow, medical, LIDAR, and astronomical photometer experiments were either entirely self-contained in data storage

and quick-look analyses or used the ADDAS system only as a back-up. The other experiments relied on ADDAS as their main data-processor; however, time and funding limitations caused experiment-ADDAS interfacing to become a major problem immediately before launch.

Lessons

Some clear lessons have been learned from the ASSESS programme, which can be used as a series of general guidelines for would-be Spacelab experimenters; the most evident are given below.

(1) Keep interfaces with Spacelab simple. This includes the mechanical, the electrical, and signal processing. The latter is particularly interesting, as Spacelab will provide a central data management system (CDMS) analogous to the ADDAS used in ASSESS II. Self-contained data gathering, processing and storage by any single experimenter not only gives him practical flexibility, and eases competition for scarce CDMS memory, but allows him to develop software at his own pace and with his own personnel during the build-up to a flight. The CDMS should be used to send highly selected data by telemetry to the ground, allowing a team to analyse problems for which the payload specialist has no time during the mission.

(2) Electromagnetic cleanliness is important. ESA set precise earthing and shielding standards for the European experimenters. These were not mandatory during ASSESS II for the US experiments, which therefore had far more problems with electromagnetic interference. Failure to adopt such standards could be costly.

(3) Experience is an effective way to learn. Those experiments which had been aboard ASSESS I gave a data yield nearly an order of magnitude greater when flown this time. This indicates that airborne missions are highly cost-effective ways for scientists to improve their potential rewards from Spacelab flights. Similar criteria apply to the management teams who feel that their direct experience on ASSESS would be valuable in organising Spacelab.

(4) For payload specialist efficiency unified control panels for groups of experiments must be built up. Diverse experiments will therefore need to be integrated, on the ground, into a single payload and also to check-out cross-talk effects, and because any payload work on board the shuttle will be costly

and complex. From a European viewpoint the build-up of a team of engineering experts at an integration site ('SPICE') within an ESA member state was shown to be practical. This centre gave a much better training scheme for payload specialists than the method used during ASSESS I, of sending payload specialists only to experimenters' home institutes for training purposes.

(5) Time and capability for *in situ* diagnosis and repairs to experiments in flight is very limited, when one payload specialist may have care of around 20 separate payload elements. A useful approach is to predict the probability of key failures, and build in redundancies as well as using modular experiment design, with replaceable key back-up units. In spite of problems, the value of the payload specialists was shown by the quantity and quality of the data obtained on ASSESS II, and clearly, using the shuttle with payload specialists will enable many more experimenters to use space platforms, with a much wider range of experiment categories than can participate in conventional satellite programmes. One result of ASSESS II is to show that the comparable complexity needed to build and operate airborne experiments, the lead times required, and the relationship to laboratory instrumentation, make aircraft programmes very suitable preparation and training for Spacelab. This is true for the experimenters, for the payload specialists, and also for the management teams. Whether ASSESS I and II have provided enough experience for NASA and ESA managements, and whether or not this is used, only time and the Spacelab programme can tell.

We thank staff members of NASA Ames Research Center, Informatics Inc., NASA Marshall Spaceflight Center, NASA Kennedy Space Center, NASA Headquarters, ESA-SPICE, DFVLR Porz-Wahn, ESA Headquarters and ESA-ESTEC who contributed to the success of this mission.

1. Vanhobst, G. *et al. Space Science Instrumentation* 1978 (in the press).
2. Rouan, D., Lena, P. J., Puget, J. L., de Boer, K. S. & Wijnbergen, J. *Astrophys. J.* **213**, L35 (1977).
3. Serra, G. *et al. Infrared Physics* (in the press).
4. Wijnbergen, J. *Astrophys. J. Lett.* (in the press).
5. Dall'Oglio, G. *et al. Phys. Rev. D* **13**, 1137 (1976).
6. Crawford, J., Rothwell, P. & Wells, N. *Nature* **257**, 650 (1975).
7. Peterson, A. W. & Kieffaber, L. *Nature* **257**, 649 (1975).
8. Peterson, A. W. & Kieffaber, L. *Nature* **242**, 321 (1973).
9. Moreels, G. & Herse, M. *Planet Space Sci.* **25**, 265 (1977).
10. Krassovsky, V. T. & Shagaev, M. V. *Planet Space Sci.* **22**, 1334 (1974).

articles

Formation of the early Earth regolith

M. D. Nussinov & A. A. Vekhov

Academy of Sciences of the USSR, Space Research Institute, 117810 Moscow, USSR

Appropriate physical conditions for the origin of life could exist on the relic regolith grains. A regolith of this type could be formed on the entire surface of the Earth under the effect of space factors in the early epoch which was characterised by a strongly rarified atmosphere. In the subsequent epoch with a denser atmosphere and suitable conditions for the first occurrence of liquid water droplets such grains had to acquire a considerable biochemical activity.

EXISTING hypotheses about the origin of life on the Earth are based either on the initial conditions of the planet or on those of space, and can be subdivided into two distinct groups. The first such group assumes that the bioevolution started in primitive water bodies¹—especially in the course of submarine volcanism², whereas the second group supports the concept of the formation of prebiological material on the surfaces of interstellar dust grains which were later transported into the Solar System and to the Earth^{3,4}. The hypotheses of the first and especially of the second group have been severely criticised^{5,6}. However, several objections can be avoided if a new hypothesis

were considered based on the following conditions necessary for 'the origin of life platforms' (developed and formulated in refs 1, 5, 7): (1) large specific area of grains (with appropriate macro- and micro-porosity); (2) immersion of grains into the water medium; (3) free energy available for biochemical reactions; (4) global character of the appropriate conditions; (5) large (in the geological sense) time span for the events; and (6) absence of powerful inhibiting factors. None of the existing hypotheses satisfies these conditions completely.

We put forward here a hypothesis with initial conditions that cannot fully be related either to the first or to the second type of hypotheses and which are 'on the border' between planetary and space conditions. According to our hypothesis, the regolith grains which formed on the early Earth's surface during an epoch of a strongly rarified atmosphere could serve as the platforms for the origin of life. We designate this type of regolith as relic or *R*-regolith. In contrast, an airless planet's regolith (such as the Moon's) will be designated *M*-regolith. We shall show that the conditions for the formation of each of these two types of regoliths are different.

Together with cosmic electromagnetic and corpuscular irradiation, meteorite bombardment is known to be one of the main exogenic factors responsible for the formation of a planet's regolith. We believe that the most interesting situation arises on a planet almost completely lacking an atmosphere ('nearly atmosphereless'). In this case the bulk of a micro-meteorite flux reaches the planet's surface with an appreciable loss in the mass and velocity of particles whose final distribution function $I_i(m_i, v_i, P)$ is transformed from the initial case $I_i(m_i, v_i, P)$ so that in the atmosphere pressure (P) range, from $P_1 \sim 10^{-7}$ atm to $P_2 \sim 10^{-4}$ atm, mostly small particles with velocities $v_i < 10^5$ cm s $^{-1}$ impinge on a planet's surface⁸. When $P > P_2$, $I_i \rightarrow 0$, extremely quickly, whereas at $P < P_1$, $(I_i - I_0) \rightarrow 0$ but much more slowly. However, the penetration of space electromagnetic and corpuscular radiations through a rarified atmosphere is significant. For instance, a CO₂ atmosphere with $P \approx 10^{-5}$ atm (which has a penetrability equivalent to a 6.5×10^{-3} cm thick CO₂ solid filter) is rather 'transparent' to the fluxes of UV and X-ray radiations and insignificantly detains charged particles with energy $E > 0.1$ MeV (ref. 9).

Hence it seems that the regolith-forming conditions on the surface of a nearly atmosphereless planet will be different to those on an atmosphereless one. Hence, a different regolith forms on each of these planets.

It is interesting to consider the problem of *R*-regolith formation in conjunction with the investigations of Meadows¹⁰ and others which establish the existence of an atmosphereless period on the Earth (after loss of the primary H₂-He atmosphere) which was followed by a period of gradual accumulation of a new atmosphere. At first, the new atmosphere consisted mainly of CO₂ and H₂O vapours—the products of mantle degassing and volcanic activity. Since then the present atmosphere has been developing. More than 4,000 Myr ago there was a crucial moment in this evolution. When the atmospheric pressures reached $P \approx 10^{-3}$ atm and the Earth's surface temperature became $T \approx 273$ K (triple point for water), the appropriate conditions for liquid H₂O occurrence were created¹¹. Before this, water desublimation resulted only in the formation of low mobile hard ice mounds and in hoarfrost. The epoch of smaller atmospheric pressures and temperatures, directly adjacent to this period, satisfied the conditions necessary for *R*-regolith formation. Therefore, we assume that by the time that water condensed, the Earth's surface was covered with an *R*-regolith which predetermined the directions of geochemical and prebiological evolutions.

Let us consider certain physical properties of *R*-regolith and, primarily, its macroporosity determined by an average grain-size distribution which, in its turn, is determined by the correlation between the rates of the two competing processes of comminution and aggregation. As for *M*-regolith, the process of comminution in *R*-regolith occurs under the impacts of micrometeorites, whereas aggregation results from cohesion

and agglutination. In a rarified atmosphere the specific energy of an impact $E_i = v_i^2/2 \approx 10^9$ erg g $^{-1}$ is sufficient only to break up the grains, and not to melt them ($E_m \approx 10^{11}$ erg g $^{-1}$ (ref. 9)). Consequently, the *R*-regolith will not contain glassy particles and glassy surfaces which are characteristic of the *M*-regolith¹². On the other hand, gas from a rarified atmosphere is immediately adsorbed on the split-off surfaces of grains which decreases drastically their aggregation due to the deterioration of the cohesive properties of grains by several orders of magnitude. Therefore, the competition between comminution and aggregation will favour the former. Consequently, the average grain-size distribution in *R*-regolith will decrease, whereas its specific surface S_R and hence its macroporosity will increase from the minimum value $S_M \approx 0.1$ m² g $^{-1}$, characteristic of *M*-regolith¹³, by one or two orders of magnitude.

Microporosity should be regarded as another peculiar feature of the *R*-regolith. Micropores result from the radiational and structural defects (tracks, dislocations, and so on) in the volume of the grains. The presence of micropores in the grains of *M*- and *R*-regoliths is due to the absorption of hard cosmic electromagnetic and corpuscular radiations. However, such defects are hardly detected in the Moon's regolith because the micropores in its grains are continuously destroyed by high-velocity micrometeorite impacts (which cause shock waves in the grain volume and melt grain surfaces) and due to sharp diurnal fluctuations of the surface temperature¹³. This has been confirmed by simulation experiments when the Moon's regolith was irradiated¹³. This situation is very similar to the conditions existing on a nearly atmosphereless planet.

The *R*-regolith micropores had to contain gas resulting from the photodecomposition of atmospheric gases (dissolved in the volume of grains) as well as rock- and mineral-forming oxides. This gas in micropores had to be in a physically adsorbed state and enriched with O₂ which accounted for about 50% of the specific weights of most rocks, minerals and some atmospheric gases.

Now, let us consider the conditions of chemical evolution which existed on the Earth's *R*-regolith during the time of liquid H₂O condensation on the planet's surface. The gas accumulated in the micropores of the grains and adsorbed on the macropores, had to escape outside by the Rebinger effect (adsorptional decrease in strength¹⁴) and by displacement desorption with water, which has the greatest energy of desorption ($E_a = 10$ –12 kcal mol $^{-1}$) relative to other common gases ($E_a = 3$ –8 kcal mol $^{-1}$ (ref. 9)). By releasing gas from the pores of *R*-regolith, water partly occupied its place and in the near-surface atmosphere, probably, initially O₂ and then other gases (CO, N₂, and so on) were evolved. However, the atmosphere remained fairly transparent to UV and other space radiations which penetrated into the pores of the *R*-regolith grains and served as a source of free energy for photo- and radiochemical processes. Thin water layers around the grains did not significantly hamper the penetration of UV and corpuscular radiations. For example, a 2 cm thick layer of water is pervious to 50% of UV radiation with $\lambda = 1,880$ Å (ref. 9), while the range of UV radiation penetration in to the grains depends on the kinds of rocks and minerals and the degree of their porosity.

The conditions (1), (2) and (3) were very favourable in the *R*-regolith for various chemical reactions to take place and different radically from those existing on the unsaturated surfaces of *M*-regolith grains. Indeed, *M*-regolith grains are cleaned by solar radiation in vacuum to an 'atomically-pure' state. Therefore, only chemisorption (that is, adsorption of a monomolecular layer) can occur on their surfaces and the grains are characterised by a low adsorption capacity¹⁵. Only simple molecules can be formed on *M*-regolith grains (for example, H₂ on the Moon's regolith¹⁶).

On the other hand, distinctive characteristics of *R*-regolith grains such as a developed system of macro- and micropores and great adsorption capacity make them resemble the interstellar dust grains despite the different origin of these

properties in the latter¹⁷. Although the precise chemical composition of the interstellar dust grains is not yet known, from the results of Hoyle, Wickramasinghe¹⁷⁻¹⁹ and others, we suppose that many have a silicate matrix, and that all are covered with a polymeric mantle resulting from chemical evolution. The interstellar dust grains in this evolution played the part of a third body providing the necessary energy and momentum balances in the chemical reactions. A similar role could be played by *R*-regolith grains. All this suggests an analogy between the conditions of chemical evolution on the interstellar dust and *R*-regolith grains. Over 40 molecular compounds, including some important for prebiological evolution such as HCN, and H₂CO, have been recorded in interstellar gas-dust clouds¹⁸. In our opinion, the conditions for complex chemical reactions to occur in the *R*-regolith of the Earth were more favourable and promoted a more rapid chemical evolution than in the interstellar gas-dust clouds owing to a higher (10^{10} – 10^{12} times) density of the ambient gas around the particles, significantly greater concentration of the *R*-regolith grains, relatively higher temperature ($T \approx 273$ K), and the presence of water medium. Indeed, if the grain surface is considered as a saturated gas then the rate ψ of chemical reactions, occurring on a grain surface surrounded by the ambient gas with the sticking coefficient equal to unity, will be proportional to n/t , where t is a characteristic time of the chemical reaction. At sufficiently high temperatures of the regolith grains (T_R), t can be found from the Arrhenius law: $t \sim \exp(-E_{act}/T_R)$, where E_{act} is the energy of activation ($E \leq 5$ eV) and T_R is the temperature in eV ($T_R \approx 0.04$ eV). The non-exponential dependence $t_d(T_d)$ is characteristic of the interstellar dust grains with their low temperature $T_d < 10$ K. According to Goldanski³, t_d may be constant $\approx 10^{-2}$ s, hence $\psi_R \gg \psi_d$.

These conditions could not only promote the formation of at least analogous and, apparently, even more complex compounds on the *R*-regolith but, what is more important, at considerably greater rates than on the interstellar dust grains. Moreover, the branching surfaces of the *R*-regolith grains acted as sites for the replication of complex linear molecular structures, whereas complex stereoregular structures were formed in the water medium in the pores (the conditions are described in ref. 20). For example, the surfaces of grains could accelerate the evolution of coacervates¹ which could be formed in the pores of the flooded *R*-regolith grains.

Hence, having compared the conditions in the interstellar dust and in *R*-regolith, one might confirm that the chemical evolution on the *R*-regolith could probably advance over a geologically short period to prebiological and biological forms, because of the fulfillment of conditions (1), (2) and (3). Conditions (4) and (5) were also present on the Earth at that time,

because the *R*-regolith covered the entire planet's surface uniformly and existed for a 'geological' period of time. The accumulation of an atmosphere created the conditions for the wind transfer of the *R*-regolith grains and, consequently, for the transportation of biological forms. At the same time, this atmosphere (as it accumulated and because oxidising) together with the increasing layer of water above the *R*-regolith grains could screen the action of the inhibiting factors. Note that unlike other hypotheses about the origin of life on planets our's does not indicate the existence of powerful inhibiting factors. For instance, having got into the water, a porous *R*-regolith grain could float for a certain time until most of its pores filled with water. That period of time might be sufficient for some prebiological reactions to occur in the grain whose products, together with the grain, could sink stepwise to the bottom of the pools to safely continue evolution there.

In conclusion, the conditions in the *R*-regolith during the epoch of water condensation satisfy the generally accepted conditions (1)–(6) necessary for the primary forms of life to have originated on Earth. It can also be supposed that such conditions have not been realised on Mars. The present status of the martian regolith is close to that of the Earth's *R*-regolith before the start of water condensation²¹.

Of course, the probability of finding the remains of the 'buried' *R*-regolith is extremely slim but its reproduction in laboratory conditions seems a possibility.

We thank Professors Shklovskii I.S. and Goldanskii V. I. and Drs I. L. Ettinger and Yu. B. Chernyak for stimulating discussions.

Received 8 August 1977; accepted 13 June 1978.

1. Oparin, A. I. *Matter, Life and Intelligence* (Nanka, Moscow, 1978) (in Russian).
2. Mukhin, L. M. *Nature* **251**, 50 (1974).
3. Goldanskii, V. I. *Nature* **269**, 583 (1977).
4. Wickramasinghe, N. C. *New Scientist*, 21 April, 119 (1977).
5. Imshenetzki, A. A. (ed.) *The Life Outside the Earth and Methods of its Discovery* (Nanka, Moscow, 1977).
6. Shklovskii, I. S. *Voprosy Phylosofii* No. 9, 80 (1976) (in Russian).
7. Sagan, C. (ed.) *C.E.T.J.* (M.I.T. Press, Cambridge, Massachusetts 1973).
8. Chernyak, Yu. B. *Nature* **273**, 497–501 (1978).
9. Kikoin, I. K. *Atomizdat* (Handbook, Physics Value Tables), Moscow (1976) (in Russian).
10. Meadows, A. J. *Planet Space Sci.* **21**, 1467–1474 (1973).
11. Rasool, S. I. & de Bergh, C. *Nature* **226**, 1037 (1970).
12. Chernyak, Yu. B. & Nussinov, M. D. *Nature* **261**, 664 (1976).
13. Holms, H. F. et al. *Earth Planet. Sci. Lett.* **1**, 1 (1973); **28**, 35 (1975); **35**, 14 (1977).
14. Rebinger, P. A. & Shchukin, E. D. *Uspekhi phys. nauk* **1**, 108 (1972) (in Russian).
15. Turner, B. E. *Astrophys. Space Sci.* **29**, 247 (1974).
16. Gibson, E. K., Jr. *Phys. Earth planet. Int.* **15**, 303 (1977).
17. Hoyle, F. & Wickramasinghe, N. C. *Nature* **264**, 45 (1976); **268**, 640 (1977); **269**, 674 (1977).
18. Zuckerman, B. *Nature* **268**, 491 (1977).
19. Greenberg, G. M. *Astrophys. J.* **189**, 181 (1974).
20. Cood, W. J. *theor. Biol.* **39**, 249–276 (1973).
21. Nussinov, M. D., Chernyak, Yu. B. & Ettinger, I. L. *Nature* (in the press).

Andesitic volcanism and crustal evolution

J. C. Eichelberger

Geosciences Division, University of California, Los Alamos Scientific Laboratory, Los Alamos, New Mexico 87545

Petrological evidence indicates that andesite is a mixture of melts from the upper mantle and lower crust. Mixing is favoured by a compressional environment and generation of rhyolitic crustal melt is favoured by continental crust. Andesitic volcanism apparently involves both addition of new material to the crust and fractionation of pre-existing crust.

A CRITICAL problem in crustal evolution is whether the voluminous andesitic volcanoes and cogenetic granodioritic batholiths are derived from the mantle, crust, or both. If these

rocks originate within the crust, they represent recycled crust. If they originate within the mantle, they represent new crust. Where they occur within the ocean basins, most commonly as the andesitic island arcs of subduction zones, they represent the generation of new continental-type material in an otherwise basaltic crustal environment.

Andesite and granodiorite do not occur alone, but belong to the basalt-andesite-dacite-rhyolite association and its plutonic equivalent, the gabbro-diorite-granodiorite-granite association. Despite the enormous range in composition and physical properties of magmas represented by these rocks, field relationships demonstrate that they are intimately related in time and space. The implication of a genetic relationship among

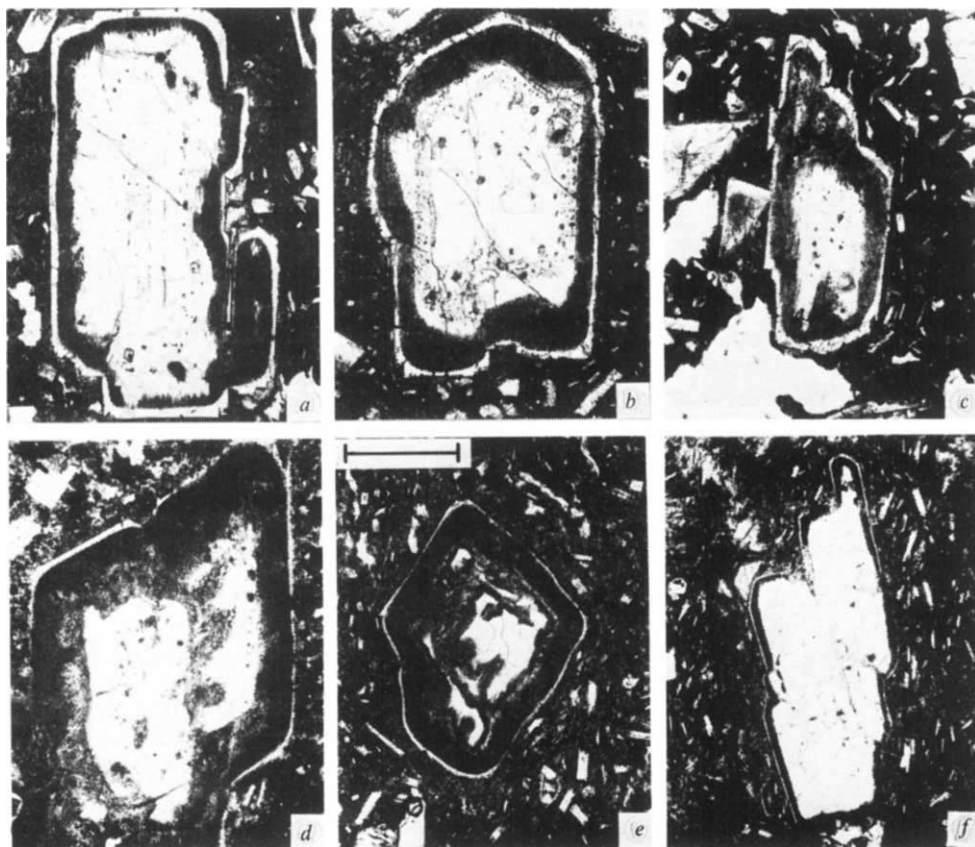


Fig. 1 Resorbed sodic plagioclase in some andesites and dacites. *a*, Glacier Peak (Washington) dacite: An 38 core, An 68 inside overgrowth, An 50 outside overgrowth. Two inclusions of rhyolitic glass (Table 2) are visible in core. *b*, Pitons du Carbet andesite (Martinique, French West Indies): An 52 core, An 75 inside overgrowth, An 56 outside of overgrowth. *c*, La Soufriere de Guadeloupe (French West Indies) andesite: An 54 core, An 83 inside overgrowth, An 57 outside overgrowth. *d*, Bill Williams Mtn dacite (San Francisco Volcanic Field, Arizona): An 23 core, An 55 inside overgrowth, An 41 outside overgrowth. *e*, Nevado Antizana dacite (Ecuador): An 25 core, An 60 inside overgrowth, An 50 outside overgrowth. *f*, Tauhara dacite (Taupo Volcanic Zone, New Zealand): An 20 core, An 61 inside overgrowth, An 46 outside overgrowth. Scale bar 0.5 mm. See Table 1 for additional data.

these magmas is strong. This relationship is a key both to the origin of these magmas and of continental crust. The purpose of this article is to discuss new data on phase assemblages in volcanic rocks (Table 1) which show a remarkable similarity among a variety of igneous complexes of this magmatic association, and which point toward mixing of basaltic and rhyolitic

magmas as a dominant mechanism by which intermediate members of the association develop. The new data include representative andesites and dacites of the Cascade Range of western United States, Taupo Volcanic Zone of New Zealand, Andes of Ecuador, and Lesser Antilles (Guadeloupe and Martinique) of the French West Indies.

Table 1 Phases in andesite and dacite derived from basaltic and rhyolitic magmas

	Volcano	Sample description	Rhyolite-derived		Basalt-derived		
			Pl*	Qz†	Pl‡	Ol§	Xenolith
ISLAND	La Soufriere Guadeloupe	Px andesite, summit dome	54*	Yes	91	82	Pl-Px
	Pitons du Carbet Martinique	Hb andesite, ~10 km N, Ft de France	52	Yes¶	92	84	Pl-Hb
	Diamant Martinique	Px andesite, near R. du Diamant	45	Yes	90	84	Pb-Px
CONTINENTAL ARC	Mt Baker Washington	Px andesite, cleaver, Eaton Glacier	46	No	82	70	Pl-Px
	Glacier Peak Washington	Hb dacite pumice, Suiattle River	38¶	Yes¶	81	81	Pl-Hb
	Mt Rainier Washington	Px andesite, Muir Camp Trail	37	No	75	70	Pl-Px
	Mt Hood Oregon	Hb andesite, White River campground	44	No	80	—	Pl-Hb
		Px andesite, Timberline Lodge	—	No	—	79	Pl-Px
	Crater Lake Oregon	Px andesite, Mazama, W. rim Crater L.	36	No	70	74	Pl-Px
	Mt Shasta California	Px andesite, Avalanche Gulch	41	No	81	78	Pl-Px
	Lassen Volcanic National Park California	Hb rhyodacite, Chaos Crags	27¶	Yes¶	92	79	Pl-Hb
		Px dacite, Lassen Pk, 1915 eruption	30¶	Yes¶	83	83	Pl-Px
		Px andesite, Prospect Pks, saddle	31¶	Yes¶	80	80	Pl-Px
CONTINENTAL ARC	Nevado Antizana Ecuador	Hb dacite, Papallacta L., recent flow	25¶	Yes¶	—	79	Pl-Hb
		Hb dacite, eroded ridge near Papallacta L.	35	Yes	71	—	Pl-Hb
	Tauhara New Zealand	Hb dacite, Burrows Quarry	20¶	Yes¶	79	82	Pl-Hb
MID CONTINENT	San Francisco	Hb dacite, Bill Williams Mt	23	Yes	—	—	Pl-Hb
	Volcanic Field, Arizona	Ol andesite, Vent 188	20¶	Yes¶	—	84	—
	Jemez Mountains	Px andesite, St Peters Dome	30	No	67	74	Pl-Px
	New Mexico	Ol basaltic andesite, TA-33, LASL	26¶	Yes¶	—	85	—

Pl, plagioclase; Px, pyroxene; Hb, hornblende; Qz, quartz; Ol, olivine

* Composition of core of resorbed plagioclase (Fig. 1) in mole % An.

† Indicates presence of quartz phenocrysts (Fig. 2).

‡ Composition of most calcic plagioclase in mole % An. Occurs as individual grains and within basaltic (cognate) xenoliths.

§ Composition of olivine in mole % Fo. Occurs as individual grains and within basaltic (cognate) xenoliths.

|| Dominant phases of basaltic (cognate) xenoliths (Fig. 3).

¶ Indicates presence of rhyolitic glass inclusions in phenocrysts.

Phase assemblages in andesite and dacite

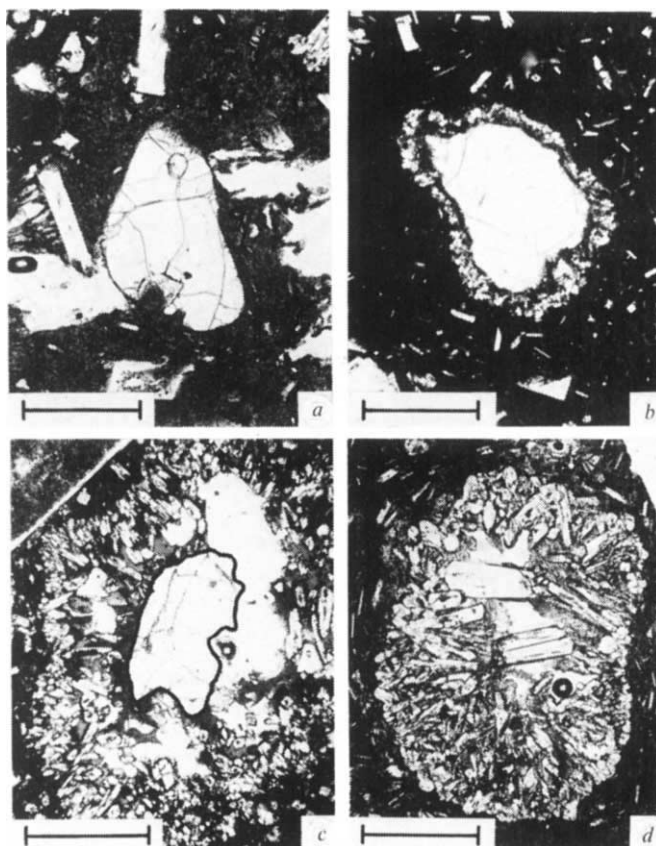
A particularly interesting feature of intermediate volcanic rocks is that they commonly contain phases totally out of chemical equilibrium¹⁻⁷ and of a type and composition appropriate to basaltic and rhyolitic magmas. In addition to the phenocrysts appropriate to magmas of intermediate bulk composition, andesite and dacite contain both anomalously calcic and sodic plagioclase, forsteritic olivine, and sometimes quartz. Sodic plagioclase occurs as cores of phenocrysts and is surrounded by a dark zone of skeletal calcic plagioclase with glass-bearing channels, which is in turn surrounded by a normally zoned clear overgrowth (Fig. 1). The width of the dark zone tends to be greater in phenocrysts in andesite than in dacite. Quartz phenocrysts in dacite commonly contain substantial embayments, indicating resorption. Quartz phenocrysts in andesite are commonly surrounded by glass and pyroxene jackets, indicating resorption and reaction (Fig. 2). Inclusions of rhyolitic glass occur both in the sodic cores of plagioclase phenocrysts and in quartz phenocrysts (Table 2). Forsteritic olivine and calcic plagioclase occur as individual grains and also within crystal-rich xenoliths and microxenoliths of distinctive texture and composition. Inclusions of basaltic glass within forsteritic olivine and calcic plagioclase have been reported by Anderson⁷.

The crystal-rich xenoliths range in size from clots of a few grains present in most thin sections (and individual crystal grains derived from them) to globules 1 m in diameter and are commonly referred to as 'cognate'. They are dominantly plagioclase + pyroxene + glass in andesite and plagioclase + hornblende + glass in dacite. Their texture (Fig. 3) consisting of euhedral to subhedral crystals with interstitial glass suggests advanced crystallisation of a magma. Plagioclase and hornblende in the xenoliths are strongly zoned, acicular, and often hollow, suggesting rapid growth¹⁵. Composition of the hornblende¹⁶ closely coincides with the composition of hornblende coexisting with melt in basaltic systems¹⁷ from 1,000 °C to 700 °C. Bulk composition of the xenoliths is also magmatic, falling at or near the basaltic end of the trend for the complex¹⁸ (Table 2). Those xenoliths which are intermediate rather than basaltic in composition contain sodic plagioclase and sometimes quartz, as described above. Large xenoliths are often concentrically zoned, with a coarse-grained basaltic core surrounded by a fine-grained basaltic or andesitic rind¹⁹. The textural relationship is clearly one of chilling of the xenolith against the host.

These features can be explained in the following way. The composition of the sodic plagioclase and quartz phenocrysts in andesite and dacite indicate that these phases originally grew in a rhyolitic magma. This conclusion is supported by the inclusions of rhyolitic glass within these phases. The composition of the calcic plagioclase and forsteritic olivine indicate that these phases grew in a basaltic magma. This conclusion is supported by the presence of these phases within the basaltic xenoliths and by the occurrence within these phases of inclusions of basaltic glass. This suggests that many andesites and dacites are hybrids derived by mixing of basaltic and rhyolitic magmas. Basaltic magma intrudes rhyolitic magma and breaks into globules. Mixing of liquids occurs at the interface between globules and host. The globules undergo rapid crystallisation to form crystal-rich basaltic xenoliths with andesitic rinds while crystalline phases in the cooler rhyolitic host undergo reaction or melting, forming embayments or pyroxene rims on quartz and dark resorbed zones on sodic plagioclase. These melting and reaction effects are greater in mixtures containing a higher proportion of basalt (andesite versus dacite)¹⁹. Renewed crystallisation following the mixing event results in growth of phases appropriate to the new intermediate bulk composition and temperature of the system, both as individual crystals and as overgrowths on basalt- and rhyolite-derived phases. Overgrowths protect the disequilibrium phases from further reaction. Note that the data presented in Tables 1 and 2 come from both oceanic and continental regions.

Other hypotheses have been advanced for the origin of these features in andesite and dacite. Sodic plagioclase and quartz in andesite have been attributed to assimilation of granitic rocks²⁰. These phases, calcic plagioclase, and forsteritic olivine have also been attributed to crystallisation of andesitic liquid at high pressure, dry in the case of sodic plagioclase²¹ and quartz^{22,23} and wet in the case of calcic plagioclase²⁴ and olivine²⁴. The clear, volcanic nature of sodic plagioclase and quartz (Figs 1 and 2) suggests that they are direct precipitates of magma. Further, the low ⁸⁷Sr/⁸⁶Sr of most andesite is inconsistent with significant contamination by older granitic rocks²⁵. Limited capacity of magmas to assimilate cold rocks and occurrence of andesite in island arcs lead to the same conclusion. This has caused many workers to favour the high pressure crystallisation hypothesis, but this leads to mutually exclusive conclusions concerning water content of the magma. Wet, high pressure conditions might produce the calcic plagioclase but not the sodic plagioclase. Such conditions might cause forsteritic olivine to precipitate but not quartz. It is also difficult to reconcile the high pressure hypothesis with evidence that there is isotopic as well as chemical disequilibrium between phenocryst and matrix in andesite and dacite^{26,27}. Basaltic xenoliths have been interpreted as cumulates^{9,11}, although they lack cumulate textures, and breakdown products of high Al amphiboles²⁸, although they range in size to several centimetres or even tens of centimetres. None of these alternative hypotheses seem to be consistent with the observations presented above, especially the inclusions of rhyolitic glass in sodic plagioclase and quartz and the textures of the xenoliths.

Fig. 2 Stages in resorption and reaction of quartz in dacite and andesite. *a*, Resorbed quartz in Glacier Peak dacite. Note two inclusions of rhyolitic glass (Table 2), one of which has leaked. Scale bar 0.5 mm. *b*, Thin, fine-grained clinopyroxene jacket on quartz in La Soufriere de Guadeloupe andesite. Scale bar 0.5 mm. *c*, Thick reaction zone of clinopyroxene + glass surrounding quartz (outlined for clarity) in Diamant andesite (Martinique). Scale bar, 2 mm. Note large resorbed plagioclase in upper left corner. *d*, End product of quartz-liquid reaction. Clinopyroxene + glass clot in Tauhara dacite. Scale bar, 2 mm. See Table 1 for additional data.



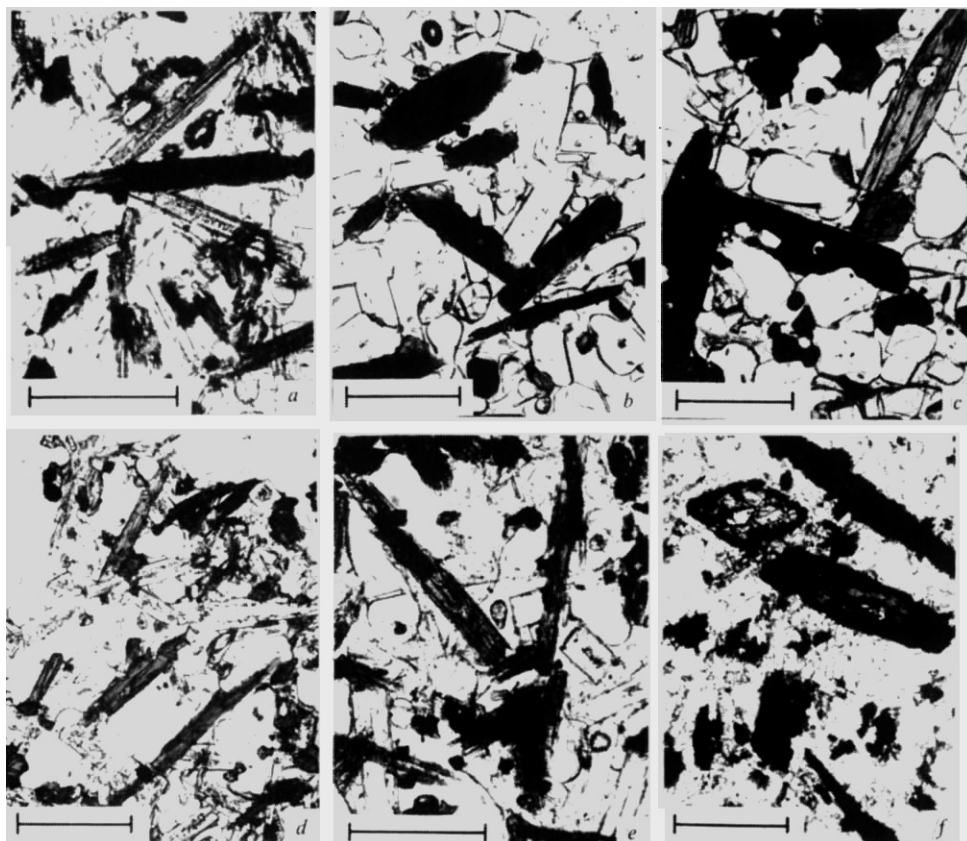


Fig. 3 Photomicrographs of plagioclase-hornblende basaltic (cognate) xenoliths in some dacites and andesites. *a*, Crater Lake dacite (Oregon). Scale bar, 0.1 mm. *b*, Mt Hood andesite (Oregon). Scale bar, 0.5 mm. *c*, Glacier Peak dacite. Scale bar, 0.5 mm. *d*, Chaos Crags dacite (Lassen Volcanic National Park, California). Scale bar, 0.5 mm. *e*, Nevado Antizana dacite. Scale bar, 0.2 mm. *f*, Pitons du Carbet andesite. Scale bar, 0.5 mm. See Table 1 for additional data.

Mixing of rhyolitic and basaltic magmas provides a simple explanation for these petrological features which are both widespread and abundant in andesites and dacites. Mixing can also account for the bulk composition of these rocks, because andesitic and dacitic compositions can be approximated (with some exceptions) as a linear combination of basaltic and rhyolitic compositions^{6,19}. Therefore it seems reasonable to formulate a general model for andesitic volcanism based on this process. Although other processes may contribute to this style of volcanism, consideration of the mixing process leads to important conclusions concerning the relationship of andesitic volcanism to crustal and tectonic setting and to crustal evolution.

Sources of parental magmas

There is general agreement that basaltic magma is a product of partial melting of ultramafic material in the mantle²⁹⁻³¹. However, widespread evidence of mixing in intermediate igneous rocks suggests that the basalt-andesite-dacite-rhyolite suite is not a liquid line of descent from basalt. If rhyolite is not a product of fractional crystallisation of more mafic magmas, it must form directly from a solid source material by a melting process. Rhyolite is too siliceous to be derived by partial melting of ultramafic mantle material or even crustal material at mantle pressures³². It does, however, represent the composition of the low melting fraction of a wide variety of crustal rocks at crustal pressures³³. This suggests that rhyolite forms by partial melting in the crust, just as basalt forms by partial melting in the mantle. The lower crust, which is the hottest part of the crust, is the most probable source region.

As the lower crust is likely to be basaltic (gabbroic) or intermediate in composition³⁴, rather than granitic, rhyolitic liquid must generally form by relatively small degrees of melting of a more mafic source. Relatively low ⁸⁷Sr/⁸⁶Sr of most rhyolites also requires derivation from a mafic (low Rb/Sr) source. The rhyolite source beneath island arcs is most likely basaltic in bulk composition because oceanic crust is formed by basaltic volcanism at the mid-ocean ridges. Experimental studies show that in hydrous conditions in the lower oceanic crust,

basaltic material will be in the form of amphibolite³⁵ and that the melt will be rhyolitic in composition up to about 20% melt fraction¹⁰. Water has a critical role because it stabilises silica-poor amphibole in the presence of silica-rich liquid. The rhyolite source is probably more silicic in the case of continental crust, because silicic lavas on continents tend to be more voluminous and more radiogenic³⁶.

Widespread evidence of mixing also implies close association of basaltic and rhyolitic magmas in time and space. Field relationships attest to this close association as well. Intermediate and rhyolitic rocks are commonly localised within regions of basaltic activity^{19,37}. Hence the rhyolitic parent of these complexes forms in a region of upwards flux of basaltic magma. Because movement of basaltic magma is a most effective means of heat transfer³⁸⁻⁴⁰, it seems likely that basaltic volcanism provides heat for generation of rhyolitic magma in the lower crust and that this is the reason for the close association of basalt with rhyolite in time and space.

Formation of hybrid magmas

The presence in andesite of phenocrysts originally grown in rhyolitic magma implies that the rhyolitic liquid segregates from its source region, coalesces into a magma body, and begins to cool and crystallise before mixing. The common occurrence of granite as large stocks or plutons, the large volume of rhyolitic extrusives, and the large collapse features associated with rhyolitic volcanism indicate that rhyolitic magma moves through the crust as large blobs or diapirs. Large volumes of megascopically homogeneous hybrids suggest that mixtures are formed in large batches. Thus it seems that much of the mixing of basaltic and rhyolitic magmas involves large (often 1 km³ or more) bodies of rhyolitic magma at levels higher and cooler than the lower crust.

Megascopic homogeneity of hybrids also implies an efficient stirring mechanism, probably convection. Application of the Rayleigh-Jeffries criteria to magmas indicates that even the most viscous rhyolitic magmas should convect in most conditions of storage in the crust⁴¹⁻⁴³. Estimated rates of flow in

plutons are of the order of 1 cm s^{-1} (ref. 43), based on the experiments of Peyches and Zortea⁴⁴ which predict reasonable rates for mantle convection. Injection of basaltic magma into an active rhyolitic pluton favours rapid and possibly turbulent convection^{39,43,45}. The system tends to homogenise itself, driven by the large temperature differences accompanying the initial heterogeneity in composition. Because convection rates are strongly dependent on size of the magma body, the stirring process is effective for pluton-sized pots but not for dispersed melt pockets in the rhyolite source region. This is consistent with the earlier observation that much of the mixing involves large batches of magma.

Field relationships shed further light on how mixing occurs. Volcanic centres which erupt intermediate and rhyolitic magmas lack basaltic vents while they are active, despite frequent basaltic activity in the surrounding region^{19,46,47}. This can be interpreted as indicating that large, active silicic magma bodies beneath these centres trap rising basaltic magma because of the lower density of the silicic magma⁴⁶. Some large volume tephra sheets from andesitic centres seem to record an early stage of mixing in which basaltic magma has been forced to sill out beneath a rhyolitic pluton⁴⁸. In the Crater Lake sheet, silicic tephra is abruptly overlain by mafic tephra. That the parent pluton was mixing, rather than differentiating, is indicated by the presence of quenched mafic material, the xenoliths described earlier, in the silicic tephra.

Statement of model

The foregoing considerations point toward a simple model for the basalt-andesite-dacite-rhyolite association (Fig. 4). Basaltic magma from the mantle rises into the crust and is emplaced at various levels as dikes and sills. In the lower crust, where the initial temperature is high, emplacement of basaltic sills causes partial melting. The resulting rhyolitic liquid gathers into plutons which rise slowly through the crust. Because the rhyolitic plutons exist in a region of upwards flux of basaltic magma, they commonly are injected with basalt and mixed by rapid convection before they crystallise or erupt. Thus basaltic volcanism causes rhyolitic magma to develop and, once formed, rhyolitic magma bodies trap rising basalt. Mixing of the two

parental magmas produces the intermediate compositions, andesite and dacite, of the association.

With this model in mind, comparison of volcanic fields of the basalt-andesite-dacite-rhyolite association in different settings suggests two important relationships: (1) generation of rhyolitic magma is favoured by continental crust and (2) mixing is favoured by a compressional tectonic environment. The first correlation was discussed above and may be due to differences in composition of the rhyolitic source. An example of how this explains variations in andesitic volcanism can be seen in the Tonga-Kermadec-New Zealand magmatic arc. Within this subduction zone, the crust of the over-riding plate varies from continental (New Zealand) to oceanic (Tonga and Kermadec). Silicic lavas of the continental portion of the arc are more voluminous, richer in K, poorer in Ca, and have higher $^{87}\text{Sr}/^{86}\text{Sr}$ values⁴⁹.

The cause of the second correlation is less clear. Extensional or rift environments, as indicated by prominent block faulting, tend to have a bimodal (basalt-rhyolite) lava distribution while intermediate compositions dominate in compressional or subduction environments⁵⁰⁻⁵². Further, changes from compression to extension in the western USA were accompanied by changes from intermediate to bimodal igneous activity^{50,51}. This suggests that extensional tectonic activity inhibits mixing, perhaps because rhyolitic plutons rise more rapidly through the crust where the crust is under tension rather than compression. Where the crustal residence time of active plutons is less, the probability of mixing is also less. Interestingly, in subduction zones, granite is more abundant among plutonic rocks than is rhyolite among volcanic rocks²⁹. Apparently there is a bias towards the eruption of mixtures. Water-rich rhyolitic plutons whose heat is not replenished by injection of basalt are more likely to crystallise at depth as granite, thereby contributing to the silicic batholiths. A further reason for the bias towards eruption of mixtures is that mixing at shallow levels can trigger eruptions by causing vapour exsolution in the magma^{43,45}.

These environmental parameters, continental or oceanic crust, compressional or extensional tectonics result in four categories which determine major characteristics of silicic igneous activity. Continental crust and compressional tectonics

Fig. 4 Model for genesis of the basalt-andesite-dacite-rhyolite association, applied to subduction and rift zone volcanism. In both cases, melting in the mantle gives rise to basaltic magma and melting in the lower crust yields rhyolitic magma. In subduction zones, intermediate lavas predominate at the surface due to mixing of basaltic and rhyolitic magmas within plutons in the mid to upper crust. In rift zones, an extensional environment permits rapid rise of rhyolitic plutons. Both basaltic and rhyolitic magmas reach the surface with little mixing. Except for differences in amount of mixing, the two cases represent a similar response of the crust to heat from the mantle. The rhyolitic component involved in volcanism is larger if the crust is more continental in character.

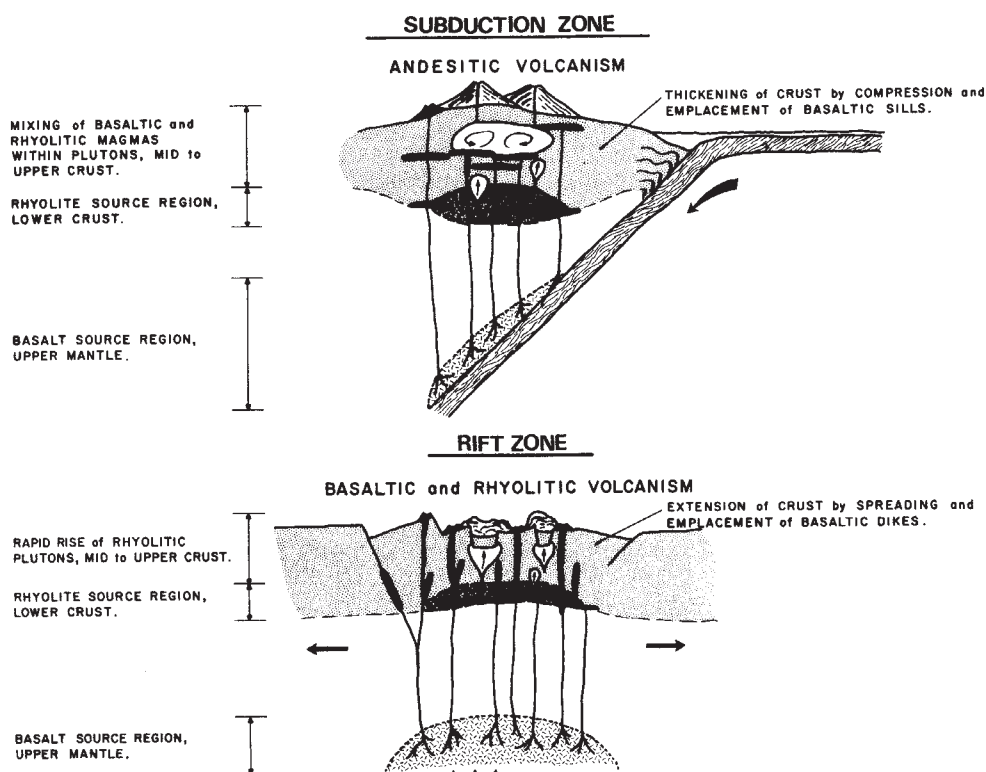


Table 2 Some observed and inferred compositions of continental margin and island arc parental magmas

		Continental margin					Island arc		
		1	2	3	4	5	6	7	8
Rhyolite	SiO ₂	74.3	75.9	75.7	74.8	73.2	73.9	73.8	74.4
	TiO ₂	0.1	0.2	0.1	0.2	0.3	0.3	0.2	0.1
	Al ₂ O ₃	13.3	12.5	11.8	12.0	13.9	12.5	13.3	16.0
	FeO	0.3	0.8	1.0	1.3	2.2	2.3	2.0	1.4
	MgO	0.0	0.2	0.1	0.3	0.4	0.4	0.4	—
	CaO	0.9	0.5	0.7	0.8	1.5	2.3	2.7	2.5
	Na ₂ O	3.8	3.9	3.7	4.0	3.9	4.0	3.9	3.7
	K ₂ O	4.8	5.1	4.2	4.2	4.2	3.2	3.3	1.8
		97.5	99.1	97.3	97.6	99.6	98.9	99.6	99.9
Coexisting plagioclase (An)		20	25		38	35	54		51
		Continental margin				Island arc			
		9	10			11	12	13	14
Basalt	SiO ₂		53.4	48.5		51.4	51.0	49.7	48.7
	TiO ₂			0.9		1.0	0.9	0.5	0.6
	Al ₂ O ₃		19.2	18.0		18.7	18.6	18.0	17.6
	FeO		7.5	8.8		8.6	8.3	10.5	11.3
	MgO		4.9	9.0		5.3	6.3	5.1	5.8
	CaO		9.8	10.7		10.7	11.0	10.0	10.4
	Na ₂ O		2.9	2.3		3.2	2.9	2.4	2.3
	K ₂ O		1.0	0.3		0.5	0.7	0.5	0.6
			98.7	98.5		99.4	99.7	96.7	97.3
Coexisting plagioclase (An)			90	84		90	91	92	85

1, Glass inclusion in resorbed sodic plagioclase, Tauhara dacite, Taupo Volcanic Zone, New Zealand (1 analysis).

2, Glass inclusions in resorbed sodic plagioclase, Nevado Antizana dacite, Andes, Ecuador (5 analyses).

3, Glass inclusions in resorbed quartz, Glacier Peak dacite, Cascade Range, Washington, U.S.A. (4 analyses).

4, Glass inclusions in resorbed sodic plagioclase, Glacier Peak dacite (10 analyses).

5, Rhyolite lava, Medicine Lake Highland, Cascade Range, California (8 analyses; ref. 8).

6, Glass inclusions in resorbed sodic plagioclase, La Soufriere andesite, Guadeloupe, Lesser Antilles (13 analyses).

7, Rhyolite lava, South Aegean Arc, Greece (5 analyses; ref. 9).

8, Melt composition, Picture Gorge Tholeiite, $T = 750^\circ\text{C}$, $P = 5$ kbar, H_2O saturated (ref. 10).

9, Basaltic (cognate) xenolith in dacite of Chaos Crags (1 analysis; ref. 11).

10, Basalt of Hat Creek Valley (1 analysis; refs 12, 13).

11, Basaltic (cognate) xenoliths in dacite of Modern Series Santorini Volcano, Greece (5 analyses; ref. 9).

12, Basalt of Main Series, Santorini Volcano (4 analyses; ref. 9).

13, Basaltic (cognate) xenoliths in andesite of Pitons du Carbet, Martinique, Lesser Antilles (2 analyses; ref. 14).

14, Basalt of Martinique (3 analyses; ref. 14).

result in a large rhyolitic component and substantial mixing. The Andes Mountains and Taupo Volcanic Zone are examples. Continental crust and extensional tectonics result in a large rhyolitic component and little mixing. The Yellowstone Plateau and East African Rift are examples. Oceanic crust and extensional tectonics result in a small rhyolitic component and little mixing. Iceland and Easter Island are examples. Finally, oceanic crust and compressional tectonics produce a small rhyolitic component and substantial mixing. Tonga and the Lesser Antilles are examples.

Evolution of the crust

In the model discussed here, andesitic volcanism involves both addition to the crust of new material from the mantle and fractionation of pre-existing crust. With time, these processes of thickening and fractionation would result in evolution from oceanic crust, thin and basaltic, to continental crust, thick and enriched in the low-melting component in its upper portion.

Andesitic island arcs, where thickened oceanic crust is trapped over a heat source, appear to represent an early stage in this evolution. Critical factors which determine whether crustal melting, and hence crustal fractionation, occurs are thickness of the crust, the presence of water, and heat transport from the mantle, either by conduction alone or augmented by movement of magma. Mixing is simply a consequence of crustal melting induced by intrusion of basaltic magma. The higher the conductive heat flow from the mantle the greater the crustal melting accompanying basaltic volcanism. Hence this process of crustal fractionation may have been more efficient earlier in the Earth's history than it is in modern island arcs.

Discussion

This view of andesitic volcanism is a departure from some current thinking on volcanic rocks, but closely resembles models for batholiths^{29,39,53,54}. As such it provides a consistent explanation for the origin of both the volcanic rocks and their

batholithic equivalents. The similarity of these volcanic and batholithic rocks extends beyond bulk composition to many of the petrological features discussed here¹⁶.

I thank R. Gooley, U. Nitsan, A. Rice, B. Crowe, and M. J. Bartholomew whose contributions are referenced here, and J. Rowley. This work was supported by the Basic Energy Sciences Division, USDOE (contract W-7405-ENG-36).

1. Fenner, C. N. *J. Geol.* **34**, 673–772 (1926).
2. Larsen, E. S., Irving, J., Gröner, F. A. & Larsen, E. S., III *Am. Mineral.* **23**, 227–257 (1938).
3. Wilcox, R. E. *Bull. Geol. Soc. Am.* **55**, 1047–1080 (1944).
4. Kuno, H. *Bull. Geol. Soc. Am.* **61**, 957–1014 (1950).
5. MacDonald, G. A. & Katsura, T. *Bull. Geol. Soc. Am.* **76**, 475–482 (1965).
6. Eichelberger, J. C. *Bull. Geol. Soc. Am.* **86**, 1381–1391 (1975).
7. Anderson, A. T., Jr *J. volcanol. geothermal Res.* **1**, 3–33 (1976).
8. Anderson, C. A. *Calif. Univ. Publ. Geol. Sci.* **25**, 347–422 (1941).
9. Nicholls, I. A. *J. Petrol.* **12**, 67–119 (1971).
10. Helz, R. T. *J. Petrol.* **17**, 139–193 (1976).
11. Williams, H. *Am. J. Sci.* **222**, 385–403 (1931).
12. Anderson, C. A. *Am. J. Sci.* **238**, 477–492 (1940).
13. Smith, A. L. & Carmichael, I. S. E. *Contrib. Mineral. Petrol.* **19**, 212–238 (1968).
14. Westercamp, D. *Bull. Volcanol.* **39**, 1–26 (1975).
15. Lofgren, G. *Am. J. Sci.* **274**, 243–273 (1974).
16. Eichelberger, J. C. & Bartholomew, M. J. *EOS* **59**, 392 (1978).
17. Helz, R. T. *J. Petrol.* **14**, 249–302 (1973).
18. Piwinski, A. J. *Neues Jahrb. Mineral. Monatsh.* **5**, 193–215 (1973).
19. Eichelberger, J. C. & Gooley, R. *Geophys. Monogr. Ser.* **20**, 57–77 (1977).
20. Sato, H. *Contrib. Mineral. Petrol.* **50**, 49–64 (1975).
21. Nash, W. P. *EOS* **54**, 507 (1973).
22. Green, T. H. & Ringwood, A. E. *Oregon Dept. Geol. Mineral. Ind. Bull. (Andesite Conference)* **65**, 21–32 (1969).
23. Marsh, B. D. & Carmichael, I. S. E. *J. geophys. Res.* **79**, 1196–1206 (1974).
24. Yoder, H. S., Jr *Oregon Dept. Geol. Mineral. Ind. Bull. (Andesite Conference)* **65**, 77–89 (1969).
25. Dickinson, W. R. *Rev. Geophys. Space Phys.* **8**, 813–864 (1970).
26. Pushkar, P. *J. geophys. Res.* **73**, 2701–2714 (1968).
27. Doe, B. R., Lipman, P. W., Hedge, L. E. & Kurasawa, H. *Contrib. Mineral. Petrol.* **21**, 142–156 (1969).
28. Stewart, D. C. *Contrib. Mineral. Petrol.* **53**, 195–204 (1975).
29. Ringwood, A. E. *Composition and Petrology of the Earth's Mantle* (McGraw-Hill, New York, 1975).
30. Carmichael, I. S. E., Turner, F. J. & Verhoogen, J. *Igneous Petrology* (McGraw-Hill, New York, 1974).
31. Yoder, H. S. *Generation of Basaltic Magma* (National Academy of Science, Washington, D.C., 1976).
32. Stern, C. R. & Wyllie, P. J. *Earth planet. Sci. Lett.* **18**, 163–167 (1973).
33. Wyllie, P. J. *Geophys. Monogr. Ser.* **14**, 279–301 (1971).
34. Ringwood, A. E. & Green, D. H. *Geophys. Monogr. Ser.* **10**, 611–626 (1966).
35. Yoder, H. S. & Tilley, C. E. *J. Petrol.* **3**, 342–532 (1962).
36. Eichelberger, J. C. *Proc. Int. Geodynamics Conf.*, Tokyo, Japan (in the press).
37. Joplin, G. A. *Geol. Mag.* **96**, 361–373 (1959).
38. Lachenbruch, A. H., Sass, J. H., Munroe, R. J. & Moses, T. H., Jr *J. geophys. Res.* **81**, 769–784 (1976).
39. Younker, L. W. & Vogel, T. A. *Can. Mineral.* **14**, 238–244 (1976).
40. Eichelberger, J. C., Gooley, R., Nitsan, U. & Rice, A. *EOS* **57**, 1024 (1976).
41. Shaw, H. R. *Am. J. Sci.* **263**, 120–152 (1965).
42. Bartlett, R. W. *Am. J. Sci.* **267**, 1067–1082 (1969).
43. Rice, A. & Eichelberger, J. C. *EOS* **57**, 1024 (1976).
44. Peckes, I. & Zortea, M. *J. geophys. Res.* **76**, 1416–1423 (1971).
45. Sparks, S. R. J., Sigurdsson, H. & Wilson, L. *Nature* **267**, 315–318 (1977).
46. Eaton, G. P. *et al. Science* **188**, 787–796 (1975).
47. Bailey, R. A., Dalrymple, G. B. & Lamphere, M. A. *J. geophys. Res.* **81**, 725–744 (1976).
48. Eichelberger, J. C. & Crowe, B. *Geol. Soc. Am. Abstr.* **10**, 104 (1978).
49. Ewart, A., Brothers, R. N. & Mateen, A. J. *volcanol. geothermal Res.* **2**, 205–250 (1977).
50. Lipman, P. W., Prostka, H. J. & Christiansen, R. L. *Phil. Trans. R. Soc.* **271**, 217–248 (1972).
51. Christiansen, R. L. & Lipman, P. W. *Phil. Trans. R. Soc.* **271**, 249–284 (1972).
52. Martin, R. F. & Piwinski, A. J. *J. geophys. Res.* **77**, 4966–4975 (1972).
53. Piwinski, A. J. & Wyllie, P. J. *J. Geol.* **76**, 205–234 (1968).
54. Brown, G. C. *Nature* **265**, 21–24 (1977).

Test of optimal sampling by foraging great tits

John R. Krebs

Edward Grey Institute of Field Ornithology, Oxford, UK

Alejandro Kacelnik

Animal Behaviour Research Group, Department of Zoology, Oxford UK

Peter Taylor

Department of Mathematics, Queens University, Kingston, Ontario, Canada

When great tits forage in an unknown environment containing two feeding places of different profitability, they first sample the two places and then exploit the more profitable one. The balance between sampling and exploitation shown by the birds is close to an optimal solution for maximising the number of food-items obtained during a feeding period.

OPTIMAL foraging models, based on the premise that animals collect food in a way which maximises their net rate of food intake, have been quite successful in predicting the decision rules used by predators in laboratory experiments^{1–9}, but so far, little work has been done on the problem of how a predator samples the environment. Sampling is an implicit necessity of optimal foraging models, which assume that the predator behaves as if it knows the availability of different prey types or patches of food. We describe here an experiment designed to test whether or not the great tit, *Parus major*, uses a maximally efficient set of rules when sampling a patchy environment.

Two possible methods of obtaining maximal food intake with a choice of foraging area

Our experimental set up is extremely simple. The naïve bird is faced with a choice of two foraging places (patches) which differ

in availability of food. The availabilities remain constant throughout a particular experiment, and the bird can only discover by exploration which of the two patches is better. Once the bird has distinguished between the quality of the two patches, it should concentrate on exploiting the better one; however, in our experiment we are concerned with how the bird makes the decision about which patch to exploit. We suggest that there are two simple types of maximising rules. On the one hand, the predator might attempt to maximise its rate of food intake at every instant in time by always foraging in the patch with the higher expected reward rate. We refer to this strategy as 'immediate maximising'. Alternatively, the predator might attempt to maximise its intake over the total foraging time and sacrifice short-term gain in order to acquire more information about the relative quality of the two patches. In this second strategy, an efficient predator has to choose the appropriate balance between exploration and exploitation. Our main aim was to test whether or not great tits follow this strategy and if so, whether they approximate an optimal balance between exploring and exploiting. The advantage of the sample-then-exploit strategy is that, in contrast to the strategy of immediate maximising, there is no risk of choosing to exploit the less profitable patch.

Simulation of the 'two-armed bandit' problem

The predator's problem of choosing the optimal balance between exploration and exploitation is similar to the classical 'two-armed bandit' problem, in which the player is faced by two

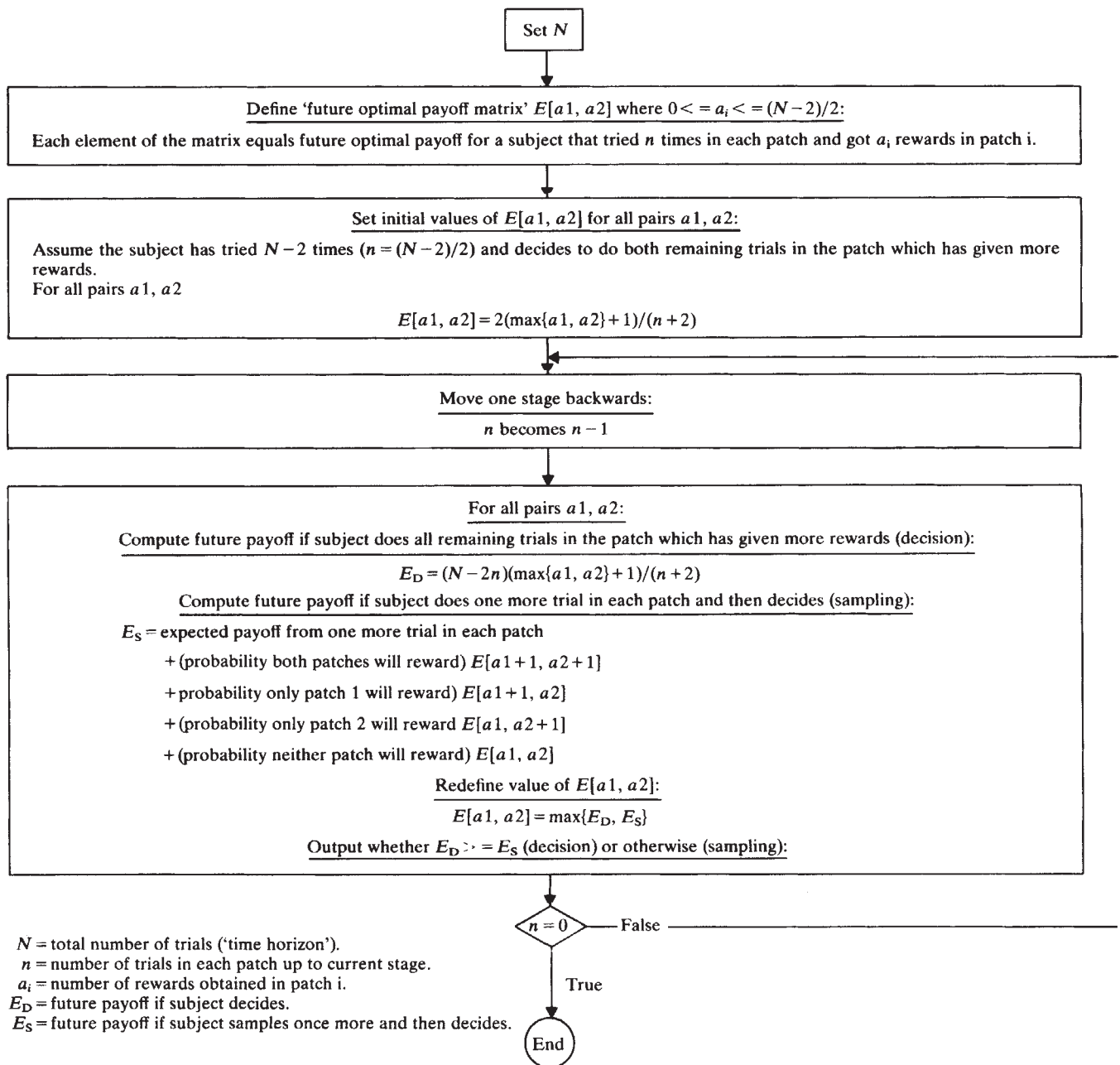


Fig. 1 'Two-armed bandit' simulation.

machines, each with a characteristic, unknown reward rate on a random schedule, and the goal is to maximise the expected number of rewards in N trials^{10,11}. This problem can be solved with a dynamic programming algorithm which works backwards from the N th trial, at which point the machine with the higher percentage rewards is chosen. The algorithm involves analysing at each stage, n , the payoff for two possibilities (exploring or exploiting) in terms of the known optimal behaviour at stage $n + 1$.

Our model is a modification of the optimal solution of the two-armed bandit problem¹². We consider that the bird has a total of N trials (the meaning of which will become apparent), and at the beginning of an experimental session it samples each patch alternately. After M trials in each patch the bird commits itself to one patch or the other, so that all strategies have the form: M trials in each patch (exploration) followed by $N - 2M$ trials on one patch (exploitation). In our model we introduced the constraint of equal sampling effort in the two patches during exploration, because this simplifies the model without significantly altering its predictions. The model predicts how the

value of M , the point at which the predator decides to exploit, changes with p_1 and p_2 , the reward probabilities in the two patches. Intuitively, one would expect M to decrease as $|p_1 - p_2|$ becomes larger, and our model predicts the exact shape of the curve.

The details of our algorithm, which uses a Bayesian approach to describe the information possessed by the bird at each stage, are as follows. At the start of the experiment, the birds' prior estimate of p (the parameter of the Bernoulli process which is the reward rate in a patch) is considered to be β -distributed with parameters (α, β) , and therefore, with mean $(\alpha + 1)/(\alpha + \beta + 2)$ and variance $(\alpha + 1)(\beta + 1)/(\alpha + \beta + 2)^2(\alpha + \beta + 3)$. The distribution has the property that after n trials with r successes, the posterior estimate of p is also β -distributed with new parameters $(\alpha + r, \beta + n - r)$. Thus, it is easy to follow the state of the bird's information about the two patches by updating α and β after each trial. In simulating the model, we used starting values of $\alpha = 0$, $\beta = 0$ for both patches. This means that the bird's initial estimate of the mean reward probability in the two patches is 0.50 with a uniform distribution between 0 and 1.0. Thus, after n

trials with a_i successes, the bird's revised estimate of the reward rate p in patch i is $(a_i + 1)/(n + 2)$. We investigated the effect of varying the prior, by calculating our predictions for the following values of α and β : $\alpha = 0, \beta = 2$; $\alpha = 0, \beta = 0$; and $\alpha = 2, \beta = 0$. The model was rather insensitive to changes in the prior, most of the changes in prediction being between 10% and 20%.

If, after n trials in each patch, the bird decides to sample once more, its expected payoff for the remainder of the foraging time is E_S , whereas if it decides to exploit, its payoff is E_D . The optimal strategy is to choose the larger (E) of these two values. E_S can be calculated in terms of the values of E at stage $n + 1$ as follows:

Let

$$\pi_i = (a_i + 1)/(n + 2)$$

Then

$$\begin{aligned} E_S(a_1, a_2, n) = & \pi_1 + \pi_2 + \pi_2 \pi_1 E(a_1 + 1, a_2 + 1, n + 1) \\ & + \pi_1(1 - \pi_2)E(a_1 + 1, a_2, n + 1) \\ & + (1 - \pi_1)\pi_2 E(a_1, a_2 + 1, n + 1) \\ & + (1 - \pi_1)(1 - \pi_2)E(a_1, a_2, n + 1) \end{aligned}$$

and

$$E_D(a_1, a_2, n) = (N - 2n) \max \{\pi_1, \pi_2\}$$

Figure 1 gives further details.

The model is simulated by starting with $2n = N$ and working backwards, generating E_S , E_D and E recursively. At each stage, the model produces an $n \times n$ matrix giving the value of $(E_D - E_S)$ for each combination of values of a_1 and a_2 . When the bird is at the last pair of trials, E_D is clearly higher than E_S , assuming there is a difference in the two reward probabilities. As the simulation works backwards, it becomes more likely that, for any particular combination of a_1 and a_2 , the payoff for sampling once more in each patch (and hence acquiring more information about the profitability of the two patches) is higher than the payoff for deciding to exploit. The exact point at which this change from sampling to decision occurs depends on the value of the total number of trials N , which in effect represents the bird's time horizon. For given values of a_1 and a_2 , the larger the value of N , the longer the bird should sample before deciding. We ran the model with a range of values of N .

The experiment

In our experiments, the two patches consisted of two identical operant feeding places at opposite ends of an indoor aviary (measuring 4.3 × 3.7 m). The feeding places consisted of a perspex disk 35.5 cm in diameter enclosed in a shallow metal box measuring 38.5 × 38.5 × 9 cm. The disk was drilled with 72 holes (diameter 0.7 cm, depth 0.4 cm) around the perimeter of its upper surface. Each of these holes contained a piece of mealworm (weight 0.07 g), and the bird had access to one hole at a time through a small horizontal window in the top of the box which enclosed the disk. To get at the next piece of food, the bird hopped on a perch next to the disk, which operated a solenoid-driven stepping cog to turn the disk through a small angle, bringing the next hole into line with the window. The reward rate was set on a pseudo-random variable ratio schedule using BRD modular logic units. This experimental set up is exactly analogous to that used in probability learning experiments¹³.

The experiments were carried out between June 1976 and February 1977 and involved testing nine wild-caught, adult great tits. Each bird was first trained to perform the operant task and then given a series of tests of different treatments consisting of various percentage reward rates in the two disks. The values we used were 50:0; 40:10; 35:15; 30:20. It is important that although these ratios always summed to 50%, we think the birds did not learn this, as between each treatment they were subjected to a series of one to four 'neutralisation' tests in which both disks offered a 7.5% reward rate. These tests were always continued until the bird showed no marked preference for one

disk over the other; we used a low reward rate during neutralisation sessions because we found this to be effective. If there was a slight preference, the less preferred disk was used as the more profitable place in the next treatment. A 'treatment' (for example, 20:30) consisted of a series of between 1 and 5 10-min tests. We had to run a series of short tests because the birds sometimes emptied one of the disks of rewards before a treatment was completed. When this happened, we interrupted the treatment for a short time while the disk was refilled. During these interruptions the bird was locked outside the test aviary without food, and in our analysis we always discounted the first 20 hops after an interruption. Any particular treatment was continued until the bird had 'decided' to exploit one disk or the other (nearly always the more profitable place).

The decision criterion was that the bird should perform more than 90% of a sequence of 100 hops on the preferred perch. This decision criterion was chosen after calculating the probability of the bird going back to the less profitable perch for a bout of 20 or more hops after different lengths of hopping bouts on the profitable perch. With the 100-hop criterion the probability of this type of reversal is nearly asymptotic and is less than 5%. Each bird was tested twice with all five treatments, the sequence of testing being randomly chosen. At the beginning of each treatment, the test bird was deprived of its normal food and excluded from the indoor aviary while the disks were being loaded with prey. Preliminary trials showed that the birds worked at a constant rate (15 hops per min) throughout a series of tests after an initial 60-min deprivation, which was the minimum period we used. The data were recorded on a computer-compatible event recorder which stored the information on magnetic tape.

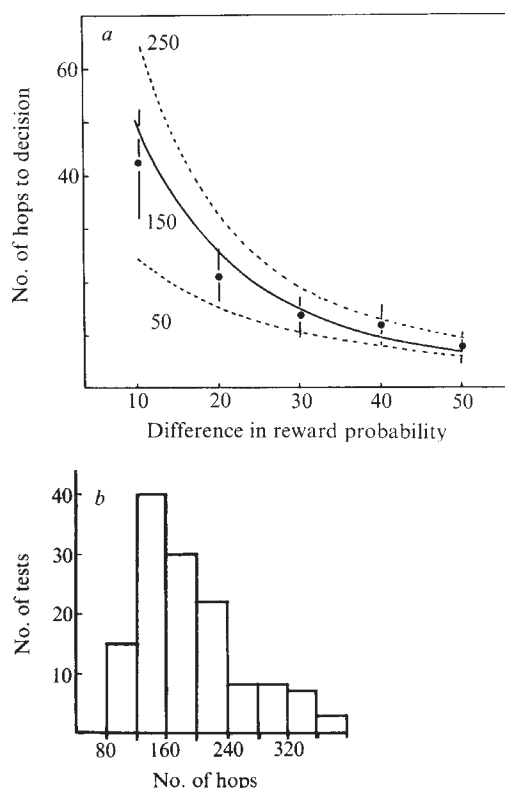


Fig. 2 *a*, The predicted (—, ---) and observed (●) number of hops before making a decision (nine birds). Three predicted curves based on the 'sample-then-exploit' model are shown, for $N = 50, 150$ and 250 . The observed means are close to 150 (continuous line) curve. Also shown are the 95% confidence intervals of the observed geometric means (we used geometric means to normalise the distribution). The x axis is the difference in reward probability between the two perches for the five experiments (for example, 50:0, difference = 50; 35:15, difference = 20) *b*, A frequency distribution of lengths of tests. The modal value is close to $N = 150$, which gave the best fit in Fig. 3*a*. $\bar{x} = 199$.

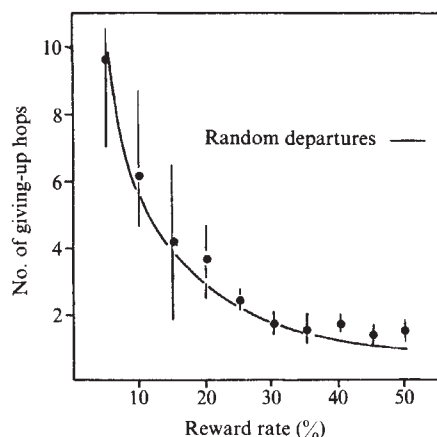


Fig. 3 The observed number of 'giving-up hops' (unsuccessful hops before a departure) plotted against reward probability. The random curve is calculated taking into account the fact that the birds tend to hop in short bouts. If p is the probability of getting a reward on a single hop and $q = (1 - p)$, then the expected number of giving up hops for bouts of 5 = $qp + 2q^2p + 3q^3p + 4q^4p + 5q^5p$. The observed points lie very slightly, but consistently above the random departure line, indicating that the birds tend to leave a perch after a run of bad luck. The observed points are $\pm$ standard error.

Treatment of results

The results of these experiments showed that the birds followed the general pattern of behaviour assumed by our version of the two-armed bandit model. There was an initial sampling period during which the birds performed on average 42% of hops on the more profitable and 58% on the less profitable perch. This proportion does not differ significantly from 50:50, and is consistent with the constraint of equal sampling in our model. The slight bias towards the less profitable perch resulted from our experimental procedure of always starting with the higher reward rate in whichever perch was used less in the preceding neutralisation test. After the decision point, the birds spent virtually all their time on the more profitable perch; on average, 95% of hops were on this perch, and the range for individual treatment means was 88–99%. Similar results, referred to as 'maximising' on concurrent variable ratio schedules, have been obtained with pigeons¹³.

In terms of our optimal decision model, we can consider a hop as being equivalent to a single trial, so that we can predict, for each combination of reward rates, how many hops on each perch the bird should do before making the decision to exploit. Figure 2a shows the observed mean decision point in each experimental treatment and the predictions of the model for three values of N (total number of hops in the experiment), 250, 150 and 50. The fit to the model with $N = 150$ is good; not only is the observed curve the same shape as predicted, but the predicted values are all fairly close to the observed means. The other values of N give curves of the same general shape, but with predicted values too high or too low. The value of N which fits the data (150) is the modal value for the number of hops performed in a test (Fig. 1b). This suggests that birds are optimising over the 'time horizon' which they experienced most often during our experiments, but we do not know if the birds would optimise with a bigger value of N if they were accustomed to longer tests. We plan to test this possibility. There is a slight difference in the modal value of test lengths for different treatments, but as the treatments were presented in a random order there is no reason to suspect that the birds could modify their behaviour according to treatment.

One way in which the birds' behaviour differed from that implied in our model is that they did not sample by doing one hop on each perch at a time, but instead they hopped in bouts of about five. (The overall mean bout length before decision was

5.00, the range 3.00–16.5.) This makes no difference in predicting the average decision point, but if the birds had hopped in rigidly determined bouts, the predicted decision point for a particular experiment would have to be a multiple of 5, which would involve reading every fifth matrix of the model's output. This tendency to hop in bouts probably explains why the observed points are above those predicted at the right-hand end of the graph where the bird should make a decision after about 10 hops.

The alternative goal we considered above was the strategy of immediate maximising, in which the bird always switches to the perch with the higher current expected reward rate. We used the β -distribution and the same starting values of α and β as above to predict, for any given experiment, at which point the birds should no longer switch back and forth if they were following this strategy. This predicted point is reached when expected payoff in the current patch never drops below that expected in the other one. The observed point was based on the decision criterion discussed above. The expected number of switches depends on the actual sequence of obtaining rewards in the first few moments of the experiment. In calculating the expected values we made the conservative assumption that the birds would continue to switch until the difference between patches in expected reward rate was greater than 5%. The observed mean values for the five treatments (50:5, 45:5, 40:10, 35:15, 30:20) were 3.3, 3.6, 3.8, 6.4 and 9.6, whereas the corresponding expected values were 0.93, 1.3, 1.36, 1.3 and 3.25. If we relax the assumption that the birds cannot discriminate differences of less than 5% in our experiments, the difference between observed and predicted values becomes even greater. Clearly, the birds switch more than expected. In other words, they are not maximising in the very short term, but instead they behave as if they acquire information to achieve a longer-term optimum.

Although the birds do not seem to switch according to the rule of immediate maximising, there is some evidence that switches between perches are related to short-term changes in reward rate. In Fig. 3 we plot the number of unsuccessful hops before a switch during the sampling period as a function of reward rate. The points lie very close to, but slightly above, the random departure curve. This random curve takes into account the fact that the birds work in bouts rather than single hops. Eight out of 10 means lie just above the random curve ($P \approx 0.05$). This shows that birds have a slight tendency to switch after a run of bad luck. In other words, they have a weak component of immediate maximising in their sampling strategy.

Discussion

The bird's behaviour approximates the prediction of a simple model based on a constrained optimal balance between exploration and exploitation, and does not fit the model of pure exploitation. The fit to the former model is close in spite of the fact that we have not yet directly incorporated a cost of switching between patches. In effect, we incorporated a small cost by using the rule that if the payoffs for sampling and exploiting were exactly equal, then the bird should exploit. In our experiment, this cost was small both in energetic terms and in time. However, in further developments of the model the cost of switching might be an important consideration.

Our comparison of the two maximising models has been in terms of predicting the details of the bird's behaviour, but an alternative approach is to compare the payoff, measured as number of rewards per test. On average, the two-armed bandit model achieves a slightly higher payoff than immediate maximising, except in very short experiments with less than five trials¹⁴, but when the reward rates are very different in the two patches (for example, 50:0) and the risk of a wrong decision is negligible, immediate maximising is just as effective. The difference between the birds' behaviour and that predicted by the two-armed bandit model, although small, is sufficient for the birds to do slightly worse than either of the two models in terms of payoff. However, the difference is less than 5% of the

predicted optimum of the two-armed bandit model (A.K. and J.K., in preparation). Thus, an analysis of payoffs enables us to say that the birds are close to the optimum but does not discriminate between the two models, unlike the analysis of the birds' behaviour.

We have presented averaged data for the nine birds we tested, but individual variation was very marked. We considered the possibility that some individuals are more cautious (do more exploring) but are less likely to make mistakes. Five of the nine individuals tested made one or more mistakes (that is, chose the poorer patch) in the 30:20 or 35:15 tests. These individuals did not make decisions after fewer hops than the other four birds. The average number of hops before a decision for all treatments was 23.9 ± 9.6 and 18.4 ± 9.8 for the fallible and infallible individuals, respectively. If, however, we examine individual treatments in which a bird made an incorrect decision, it is apparent that these tend to be cases in which the bird did little or no sampling. Examining the 18 '20:30' treatments (nine birds, two treatments each), only three out of the 13 in which a correct decision was made involved no sampling, whereas two out of the five in which the bird made a mistake were tests in which there was no sampling period at the beginning.

Although the birds in our experiments behaved on average as if they were calculating an optimal balance between exploration and exploitation, we suggest that they use some simple rule to

approximate the optimum. We have simulated various potential rules including 'matching' future responses to rewards obtained so far, an empirically derived rule applying concurrent choice experiments¹⁵. The extent to which these proximate rules mimic the behaviour of the birds and match the total payoff achieved will be discussed elsewhere (A.K. and J.K., in preparation).

We thank the SRC for financial support, Mike Cullen and John Gittins for valuable advice, and Nick Styles, Janet Lomer and Ruth Ashcroft for help with the experiments. Ted Fleming and Nick Davies commented on the manuscript.

Received 20 March; accepted 18 July 1978.

1. Krebs, J. R. & Davis, N. B. (eds) *Behavioural Ecology: An Evolutionary Approach* (Blackwell, Oxford, 1978).
2. Krebs, J. R., Ryan, J. C. & Charnov, E. L. *Anim. Behav.* **22**, 953-964 (1974).
3. Krebs, J. R., Erichsen, J. T., Webber, M. I. & Charnov, E. L. *Anim. Behav.* **25**, 30-38 (1977).
4. Cowie, R. J. *Nature* **268**, 137-139 (1977).
5. Goss-Custard, J. D. *Anim. Behav.* **25**, 10-29 (1977).
6. Cook, R. M. & Hubbard, S. F. *J. Anim. Ecol.* **46**, 115-125 (1977).
7. Charnov, E. L. *Theor. Populat. Biol.* **9**, 129-136 (1976).
8. Charnov, E. L. *Am. Nat.* **110**, 141-151 (1976).
9. Werner, E. E. & Hall, D. J. *Ecology* **55**, 1042-1052 (1974).
10. Bellman, R. *Sankhyā* **16**, 221-229 (1956).
11. Barnett, V. *Comparative Statistical Inference* Wiley, New York, (1975).
12. Lindley, D. V. & Barnett, B. N. *Biometrika* **52**, 507-532 (1965).
13. Mackintosh, N. J. *The Psychology of Animal Learning* (Academic, New York, 1974).
14. Jones, P. W. *Biometrika* **62**, 523-524 (1975).
15. Herrnstein, R. J. & Loveland, D. H. *J. exp. Analysis Behav.* **24**, 107-116 (1975).

Electrophoretic movement and localisation of acetylcholine receptors in the embryonic muscle cell membrane

Norman Orida & Mu-ming Poo

Department of Physiology, California College of Medicine, University of California, Irvine, California 92717

A steady electric field of 30 mV across a single embryonic muscle cell produces accumulation of acetylcholine receptors toward one pole of the cell within 1 h. The movement is electrophoretic in nature and the accumulation results in the formation of stable, metabolically independent receptor aggregates.

MANY cell membranes are fluid in structure¹, and proteins and lipids are free to undergo long-range movement in the plane of the membrane²⁻⁵. Under many circumstances, however, mobile membrane components may be localised and concentrated at specific regions of the cell membrane. The mechanism that modifies and stabilises their topographic distribution in the fluid matrix is poorly understood. One notable example is the acetylcholine (ACh) receptors at the neuromuscular junction. Embryonic muscle membrane before innervation is sensitive to ACh over the entire surface of the muscle^{6,7}. After synaptogenesis, the ACh sensitive site is essentially confined to the junctional region of the muscle membrane^{6,8,9}. Embryonic muscle cell membranes are fluid^{10,11}, and redistribution of ACh receptors during the course of innervation has been demonstrated^{12,13}. We present here our study on electric field induced movement and localisation of the ACh receptor in muscle cell membranes. We found that when a steady electric field of 10 V cm^{-1} (corresponding to a potential difference of 30 mV across a cell $30 \mu\text{m}$ in diameter) was applied along the surface of spherical *Xenopus* muscle cells, the ACh receptors

rapidly accumulated towards one pole of the cell. This accumulation of receptors is consistent with the electrophoretic redistribution of mobile, charged receptors in the plane of the cell membrane. The redistribution is independent of metabolic energy, blocked by the preincubation of the cells with concanavalin A (Con A) and reversed in direction after surface charge modification with neuraminidase. Moreover, the accumulation of the ACh receptors leads to the immediate formation of stable receptor aggregates which persist against back diffusion, resist treatment with cytoskeletal disrupting agents, and require no metabolic energy supply for their stability.

Ionophoretic mapping of ACh receptors

Single embryonic muscle cells, obtained by dissociating the neural tubes of *Xenopus laevis* embryos (stage 18-19, ref. 14), were plated as a monolayer on clean glass culture chambers and were used for experiments after 2.5-3.5 d in culture. Culture medium contained 85% Steinberg's saline¹⁵ (58 mM NaCl, 0.7 mM KCl, 0.4 mM $\text{Ca}(\text{NO}_3)_2$, 0.1 mM MgSO_4 and 4.6 mM Tris), 10% Leibovitz medium (L-15, Gibco), and 5% fetal calf serum (Gibco). Within 2.5 d of culture, embryonic muscle cells of two distinct morphologies were observed, extended spindles and spheres. All experiments were carried out on the isolated, spherical mononuclear cells (diameter $35 \pm 5 \mu\text{m}$) that adhered tightly to the culture substratum (Fig. 1a). Details of electrophoresis methods and apparatus have been reported elsewhere¹¹. Briefly, an electric current was applied to the thin, rectangular culture chamber of defined geometry ($6.0 \times 1.0 \times 0.02 \text{ cm}$) containing culture medium. A current of 1.5 mA

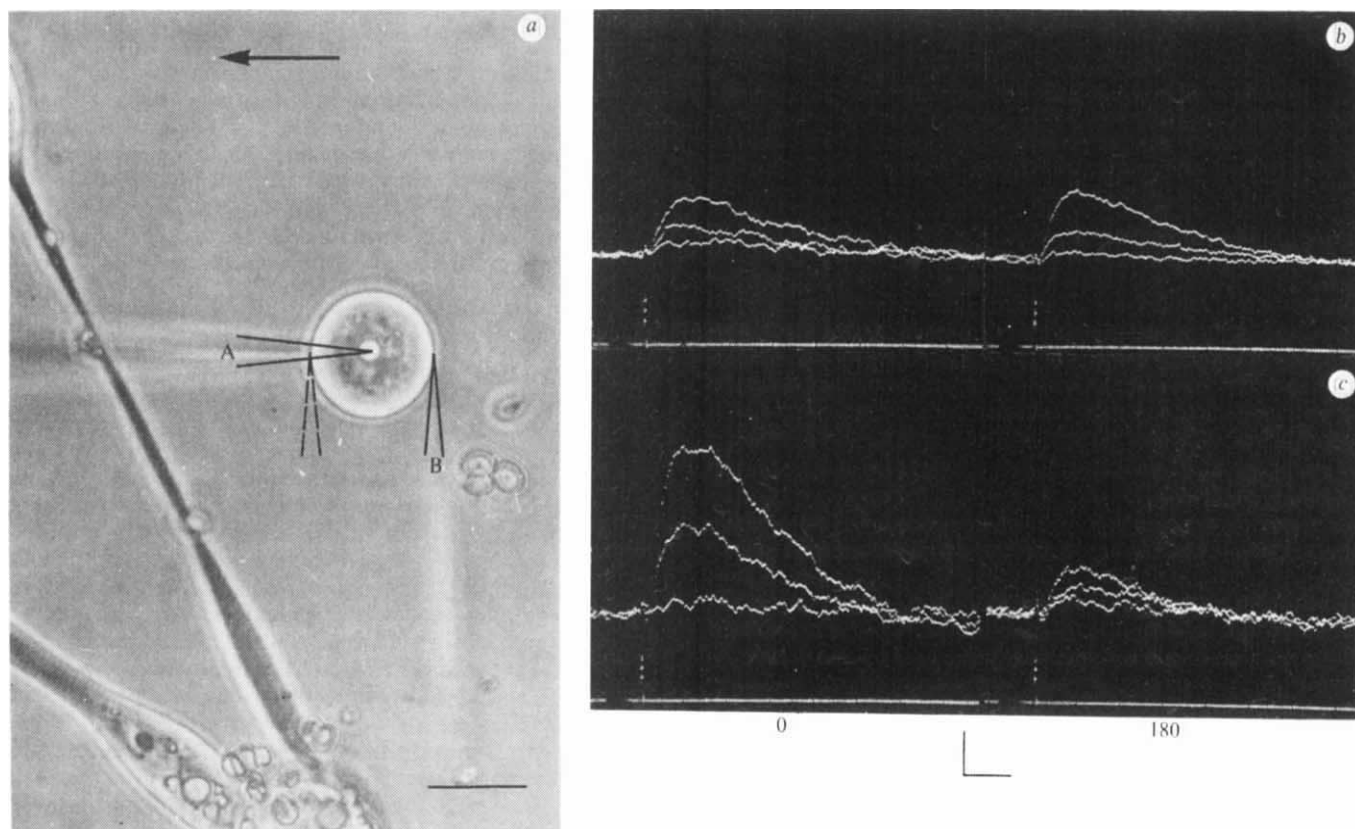


Fig. 1 *a*, Monolayer culture of *Xenopus* embryonic muscle cells after 2.5 d of growth. Also shown are recording micropipette A and ACh micropipette B used for iontophoresing ACh along the perimeter of spherical cells. Intracellular recording was done using glass microelectrodes filled with 3 M potassium acetate (resistances 120–200 M Ω). Microelectrodes for iontophoretic application of ACh were filled with 3 M acetylcholine chloride (Sigma), had resistances from 200 to 350 M Ω , and required braking currents between 1.0 and 2.0 nA to prevent ACh leakage and depolarisation of the cells. Current pulses were delivered through a preamplifier (Getting microelectrode) to the ACh pipette. Compensation for pipette capacitance by the negative capacity preamplifier allowed accurate estimates of charge delivered in the 'squared-off' current pulses. Currents ejected from the ACh pipettes were measured with a current-to-voltage converter which kept the bath at virtual ground. All electrophysiological recording was done in 100% Steinberg's saline supplemented with 10 mM CaCl₂. Cells were observed with phase contrast optics at magnification $\times 350$. Bar = 30 μ m. Arrow indicates direction of the field. *b*, Potentials recorded in response to iontophoretically applied ACh during sensitivity mapping of a representative control cell (not exposed to the field). Symmetrical, graded potentials (upper trace) were recorded for identical iontophoretic current pulses (lower trace) which were delivered to the two poles of the cell (0° and 180°). Horizontal bar, 10 ms; vertical bar, 5 mV voltage, 5 nA current; $E_m = -104$ mV. *c*, Asymmetrical response from another cell in the same culture as *b* after exposure to a field of 10 V cm⁻¹ for 1.5 h. Again, identical current pulses were delivered from the same ACh pipette to the two poles. Scales as in *b*; $E_m = -84$ mV.

produced the standard field of 10 V cm⁻¹ along the culture substratum. The power dissipation by this current produced a small temperature rise (<1.5 °C) in the chamber, as measured with a thermistor. Exposure to the standard field did not induce preferential locomotion of these muscle cells, or cause any rotation of the cells, as determined by the maintained position of yolk granules occasionally attached to the cell surface. All experiments were carried out at room temperature (20–22 °C).

For mapping the topographic distribution of ACh receptors, micro-iontophoresis methods were used. In a typical mapping experiment, a cell was first penetrated with the recording electrode. Successful penetrations resulted in resting membrane potentials of from -80 to -120 mV ($\bar{x} = -101 \pm 10$ mV s.d.), which remained stable up to 30 min. These resting potentials agree well with those of another study on the same cells¹⁶. To map the ACh sensitivity, the ACh electrode was positioned by approaching the cell perimeter until a slight dimple appeared at the desired site of mapping. The ACh electrode was then backed up until the dimple disappeared. In this condition, current pulses of 0.5 ms duration and graded amplitudes delivered to the ACh pipette produced graded ACh potentials with rise time between 6 and 15 ms (average ≈ 9 ms). Peak ACh depolarisation was plotted against the number of coulombs ejected from the ACh pipette. The linear portion of these curves was used to express ACh sensitivity in V nC⁻¹ (ref. 17).

In control cultures not exposed to the field, equal sensitivity was observed at the two poles of the cells (0° and 180° with respect to the chamber and field axis). For 72 cells mapped in 11 separate cultures, the sensitivities were 9.45 ± 0.65 and 9.44 ± 0.63 V nC⁻¹ at 0° and 180° poles, respectively (Row 1, Table 1, Fig. 1*b*, 2*a*). These sensitivity values are higher than those in other reports^{18,19}, probably through a combination of the high input resistance of these cells (150 M Ω , unpublished observations) and the large driving force due to highly negative resting potentials. The measured sensitivities were quite reproducible for each cell. As an example, for 17 determinations at the 180° pole of a control cell made through repeated repositioning, the V nC⁻¹ values ranged from 5.0 to 8.3 and averaged 6.75 ± 0.42 (95% confidence levels).

Redistribution of ACh sensitivity by electric field

When an electric field of 10 V cm⁻¹ was applied for 1.5 h to the muscle cells and the ACh sensitivity mapped after removal of the field, we found that ACh sensitivity over the cell surface became grossly asymmetrical. For 42 cells mapped in five separate cultures, the sensitivities on the cathodal side of the cells facing the negative pole of the field (0°) averaged 18.9 ± 2.3 V nC⁻¹, whereas that on the anodal side of the cells (180°)

averaged $6.3 \pm 1.0 \text{ V nC}^{-1}$ (row 2, Table 1). Figure 1c shows representative asymmetrical responses of a cell after 1.5 h exposure to the field, and Fig. 2 plots peak depolarisation against charge ejected through the ACh pipette for two cells from the same culture. The sensitivities on the two poles of these cells were mapped before (cell *a*) and after (cell *b*) exposure to the field for 1.5 h. For the cell exposed to the field, the slope of the curve increased at the cathodal pole, and decreased at the anodal pole. In several experiments, we also mapped the sensitivity along the perimeter of the cell from 0° to 180° . It was consistently found that the positions with increased sensitivity were always centred around the 0° pole, whereas the sensitivities of all other positions decreased slightly from the control level. Figure 3 shows perimeter mappings for a control cell and three cells from three separate cultures exposed to the standard 10 V cm^{-1} field for 45 min, 1.5 h and 3 h, respectively.

To evaluate the degree of asymmetry in ACh sensitivity induced by the electric field, we defined an asymmetry index with the formula described in Table 1. This formula gives an index of 0 for a cell showing equal sensitivities on the cathodal and anodal pole of the cell, an index of 1 when the sensitivity is entirely confined to the cathodal pole, and an index of 0.5 when the sensitivity is threefold higher at 0° than that at 180° . By plotting mapping data for a series of cultures exposed to different durations of the same field, we found that the redistribution of ACh sensitivity reached a plateau asymmetry index of about 0.5 (Fig. 4).

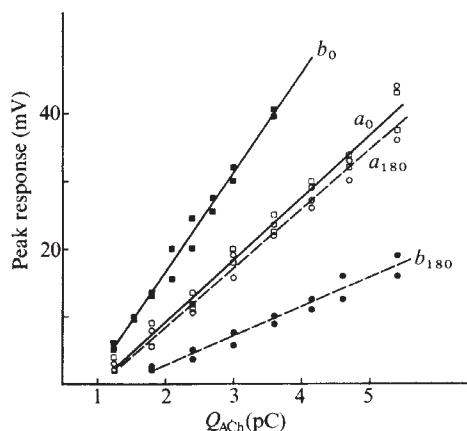


Fig. 2 Dose-response curve of the muscle cell membrane to applied ACh. Curves were obtained by plotting peak depolarisation produced by the picocoulombs of charge ejected from the ACh pipette. Slopes of lines drawn by eye were used as a measurement of sensitivity. Cells *a* and *b* were from the same culture and were mapped with the same ACh pipette before (*a*) and after (*b*) exposure to a field of 10 V cm^{-1} for 1.5 h. Sensitivities at 0° and 180° were 9.3 and 8.8 V nC^{-1} , respectively, for cell *a*, and in contrast, 14.5 and 4.3 V nC^{-1} , respectively for cell *b*. E_m during mapping ranged between -100 to -104 and -94 to -98 mV for *a* and *b*, respectively.

The asymmetric redistribution of ACh sensitivity was not due to the redistribution of acetylcholinesterase, the hydrolysing enzyme for ACh, over the cell surface. When iontophoretic mapping was carried out using a pipette filled with 3 M carbachol (Sigma), a non-hydrolysable ACh analogue, similar asymmetrical redistribution of sensitivity was found (row 4, Table 1). It is interesting, however that the absolute sensitivity of these cells towards carbachol is about fivefold lower than that towards ACh in identical iontophoretic conditions.

The redistribution of ACh sensitivity was a process independent of cell metabolism. In three separate experiments, cultures were preincubated in saline containing 10^{-2} M sodium azide, 10^{-3} M dinitrophenol and 10^{-3} M sodium fluoride for 45 min. This treatment blocks the contraction response of these muscle cells to bath-applied ACh (1 mM), indicating a

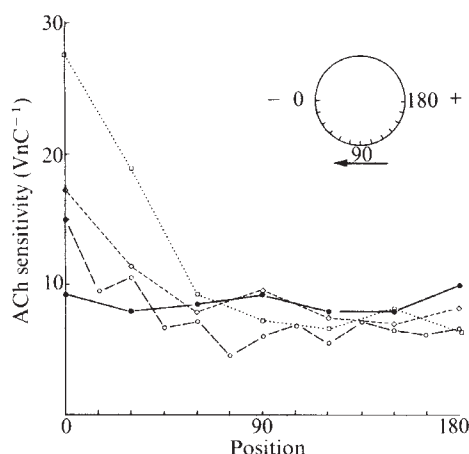


Fig. 3 Mapping of ACh sensitivity at various points along the perimeter of spherical muscle cells (inset). Four cells from four different cultures were exposed to a field of 10 V cm^{-1} for 0 min (●—●), 45 min (○—○), 1.5 h (◇—◇) and 3 h (□—□). Arrow indicated direction of field.

depletion of metabolic energy. Cells so treated were then exposed to the standard field for 1.5 h, still in the presence of the metabolic inhibitors. Post-field mapping indicated that although the absolute sensitivities decreased on both sides of the cells (due to progressive loss of the membrane potentials), similar redistribution of ACh sensitivity still occurred (compare rows 2 and 5, Table 1). Consistent with the finding that the observed redistribution is a passive process driven by the external field, we found that cytoskeletal elements do not seem to be involved in the redistribution process. Preincubation of the cultures in saline containing cytochalasin B (Sigma, $10 \mu\text{g ml}^{-1}$) and colchicine (Sigma, $20 \mu\text{M}$) for 1 h does not affect the redistribution of ACh sensitivity subsequently induced by the field (row 6, Table 1). Such treatment is known to block active capping processes of surface receptors in other cell types²⁰.

The effect of Con A and neuraminidase

The above observations indicated that the redistribution of ACh sensitivity is consistent with a passive electrophoretic redistribution of ACh receptors in the plane of the cell membrane. If there was indeed a redistribution of the pre-existing ACh receptors, immobilisation of ACh receptors

Fig. 4 Rise of asymmetry index with increased duration in an electric field (10 V cm^{-1}). Data were from many experiments grouped for mappings done within 4 h of removal from field. Error bars represent 95% confidence levels. The figures indicate the number of cells mapped for each duration. Data with dashed error bars indicate cultures which were preincubated with Con A for 10 min before the field was applied.

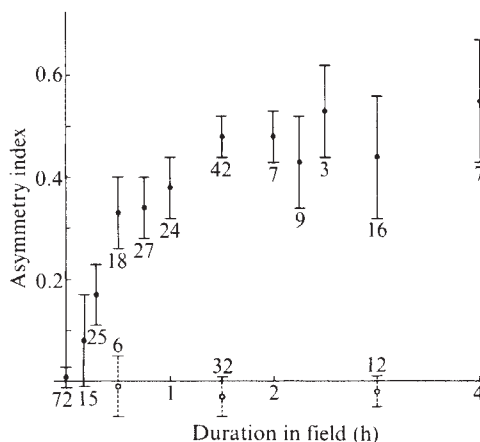


Table 1 Redistribution of ACh sensitivity induced by electric field

Duration of field exposure (10 V cm ⁻¹)	No. of cells mapped*	ACh sensitivity† (V nC ⁻¹)		Asymmetry index‡
		At 0° (cathodal pole)	At 180° (anodal pole)	
1. Control (no treatment)	72 (11)	9.45 ± 0.65	9.44 ± 0.63	0.01 ± 0.02
2. 1.5 h	42 (5)	18.9 ± 2.3	6.3 ± 1.0	0.48 ± 0.04
3. 30 min	18 (3)	18.4 ± 3.8	9.1 ± 1.5	0.33 ± 0.07
4. 1.5 h (carbachol pipette)	17 (1)	4.0 ± 0.9	1.3 ± 0.4	0.49 ± 0.08
5. 1.5 h (metabolic inhibition)	14 (3)	15.6 ± 5.0	6.4 ± 1.5	0.37 ± 0.11
6. 1.5 h (treatment with cytochalasin B and colchicine)	15 (1)	10.6 ± 2.0	4.2 ± 0.6	0.42 ± 0.06
7. 1.5 h (preincubation with Con A)	19 (2)	7.3 ± 1.4	7.9 ± 1.2	-0.04 ± 0.08
8. 1.5 h (preincubation with neuraminidase)	22 (3)	6.5 ± 1.5	14.1 ± 2.4	-0.37 ± 0.08

*All isolated spherical cells mapped successfully (membrane potential $|E_m| \geq -80$ mV, $\Delta E_m \leq 10$ mV during mapping) were included. Data were pooled from separate cultures (number of cultures shown in parentheses). As there was no relaxation of the asymmetry after the removal of field (see last section and Fig. 5), data were pooled regardless of the exact time the mapping was carried out.

†Sensitivity presented as (average) ± 95% c.l.

‡Asymmetry index = Sensitivity at 0° - Sensitivity at 180° / Sensitivity at 0° + Sensitivity at 180°, was calculated for each cell before averaging, and presented as (average) ± 95% c.l.

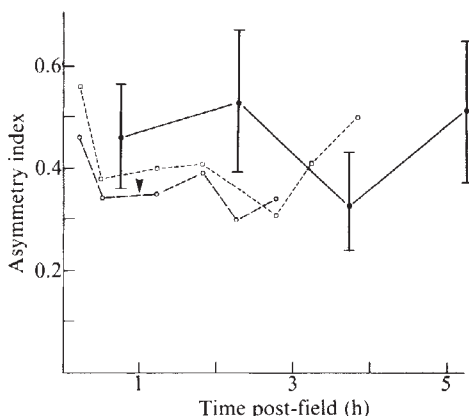
before the application of the field should prevent the redistribution of ACh sensitivity. The following experiments using Con A provide such a control. Con A is a tetravalent lectin molecule that specifically binds to the α -D-mannose or α -D-glucose residues of cell surface glycoproteins and glycolipids²⁰, and is known to crosslink and immobilise surface receptors²¹. In control cultures not exposed to the field, incubation with Con A (Sigma, grade III, 150 μ g ml⁻¹) for 10 min did not greatly affect the absolute ACh sensitivity of the muscle cells. The distribution of ACh sensitivity was essentially uniform along the whole perimeter of the cells. For the cells preincubated with Con A then exposed to the standard field for various times, no significant asymmetry in the sensitivity was produced at the 0° and 180° poles (Fig. 4, open circles, and row 7, Table 1). We have shown previously¹¹ that preincubation with Con A blocks the electrophoretic accumulation of the Con A receptors, as indicated by post-field fluorescent Con A ring staining. As the ACh receptor is probably a subset of Con A receptors²², the present finding is consistent with the idea that immobilisation through Con A binding prevents electrophoretic movement of the ACh receptors.

The surprising effect of neuraminidase treatment on the redistribution of ACh sensitivity provided additional evidence that the redistribution is electrophoretic in nature. Neuraminidase removes sialic acid from the cell surface, reducing the negative charges on membrane constituents²³. We preincubated our culture with neuraminidase (Sigma, grade VI, 0.1 unit ml⁻¹) for 1 h at pH 6.6. The enzyme-containing saline was then substituted with culture medium and the field of 10 V cm⁻¹ was applied. Post-field mapping showed that the redistribution of ACh sensitivity reversed its polarity for all cells mapped (22 cells in three separate experiments; row 8, Table 1), as compared to cells not treated with neuraminidase. The sensitivity at the anodal pole increased greatly whereas that at the cathodal pole decreased. The reversal in the direction of redistribution is probably due to the modification of charged groups on the cell surface by the enzyme. The fact that ACh sensitivity increased on the cathodal side of the cells not treated with neuraminidase suggests that ACh receptors are positively or less negatively charged than other mobile membrane components. It is likely that ACh receptors are, in fact, negatively charged but contain fewer sialic acid moieties than other more negatively charged components. After neuraminidase treatment, they become among the most negatively charged components, and hence, accumulate at the anodal pole. It should be noted, however, that a full understanding of the electrophoretic redistribution of various membrane components still requires more information on the interaction between the charged groups on the cell surface and on the effect of various counter-ions on the charged moieties.

Stability of receptor aggregates

An interesting finding in the present study is that field-induced asymmetrical ACh sensitivity always persists after removal of the field. Figure 5 shows the asymmetry indices for cells mapped at various times after removal from a 1.5 h standard field. Previously¹¹, it was shown that field-induced accumulation of the most mobile population of Con A receptors in these muscle cells undergoes relaxation after removal from the field. Rapid back diffusion results in the recovery of uniform distribution within 30 min. In our present study, for cultures exposed to a field of 10 V cm⁻¹ for 1.5 h, the asymmetry indices persisted for as long as the mapping was carried out (up to 5 h). The stability of the field-induced receptor aggregates is not dependent on cell metabolism. Figure 5 (open circles) shows a series of mapping data obtained from cultures exposed to the field for 1.5 h, but in which the regular saline was replaced (at arrow) with saline containing metabolic inhibitors (10⁻² M sodium azide, 10⁻³ M dinitrophenol). For as long as the cells remained viable under this treatment, we found no significant change in the stability of the asymmetry. Drugs disrupting cytoskeletal structure also produced no effect on the stability.

Fig. 5 Stability of asymmetrical ACh sensitivity after removal from the electric field. Data from 10 separate cultures were grouped into 90-min bins and averaged for cells exposed to the field for 1.5 h without further pharmacological treatment (●—●). Data were grouped into 30-min bins and averaged for cells either exposed to the field for 1.5 h, and then treated with sodium azide and DNP (at arrow) for the remainder of the experiment (○—○) or treated with cytochalasin B and colchicine (□—□) during the entire experiment (1 h preincubation, 1.5 h field, and post-field mapping). Each data point, on average, represents seven independent measurements. For clarity, error bars (±95% c.l.) are included only for the 90-min bins.



Asymmetry persisted for up to 4 h in the presence of $10 \mu\text{g ml}^{-1}$ of cytochalasin B and $20 \mu\text{M}$ of colchicine in the saline (Fig. 5, open squares).

Although stable against back diffusion, the field-induced ACh receptor aggregates were still electrophoretically mobile. When a field of the same strength, but different polarity was applied to a culture previously exposed to the standard field, a second accumulation of the ACh receptors occurred at the new cathodal poles of the cells. The sensitivity at the original cathodal pole declined to the control level (unpublished data). Our observations on the accumulation of ACh receptors by the field, the stability of aggregates against back diffusion and their mobility under subsequently imposed fields of different polarity have now all been confirmed using fluorescently labelled α -bungarotoxin²⁴, a venom protein known to bind specifically to ACh receptors.

In conclusion, we have demonstrated the electrophoretic movement and localisation of a specific membrane protein/ionic channel in the plane of the cell membrane. The technique of *in situ* electrophoresis reported here may prove useful in perturbing the receptor topography on the cell surface and in dissecting the dynamics of receptor interactions. Whether endogenous electric fields within developing tissue are involved in the localisation of cell surface receptors remains to be elucidated.

This work is supported by a grant from the USNSF to M-m.P.

Received 17 March; accepted 14 June 1978.

1. Singer, S. J. & Nicolson, G. L. *Science* **175**, 720–731 (1972).
2. Edidin, M. *Ann. Rev. Biophys.* **3**, 179–201 (1974).
3. Lee, A. G. *Prog. Biophys. molec. Biol.* **29**, 3–56 (1975).
4. Poo, M-m. & Cone, R. A. *Nature* **247**, 438–441 (1974).
5. Poo, M-m. & Robinson, K. R. *Nature* **265**, 602–605 (1977).
6. Diamond, J. & Miledi, R. *J. Physiol., Lond.* **162**, 393–408 (1962).
7. Sytkowski, A. J., Vogel, Z. & Nirenberg, M. W. *Proc. natn. Acad. Sci. U.S.A.* **70**, 270–274 (1973).
8. Cohen, S. A. & Fischbach, G. D. *Dev. Biol.* **59**, 24–38 (1977).
9. Kidokoro, Y., Heinemann, S., Schubert, D., Brandt, B. L. & Klier, F. G. *Cold Spring Harb. Symp. quant. Biol.* **40**, 373–388 (1976).
10. Edidin, M. & Fambrough, D. J. *Cell Biol.* **57**, 27–37 (1973).
11. Poo, M-m, Poo, W-j. H. & Lam, J. J. *Cell Biol.* **76**, 483–501 (1978).
12. Anderson, M. J. & Cohen, M. W. *J. Physiol., Lond.* **268**, 757–773 (1977).
13. Fischbach, G. D., Berg, D. K., Cohen, S. A. & Frank, E. *Cold Spring Harb. Symp. quant. Biol.* **40**, 347–357 (1976).
14. Nieuwkoop, P. D. & Faber, J. *Normal Table of Xenopus laevis* (North-Holland, Amsterdam, 1962).
15. Jones, K. W. & Elsdale, T. R. *J. Embryol. exp. Morph.* **11**, 135–154 (1963).
16. Spitzer, N. C. *J. Physiol., Lond.* **255**, 105–135 (1976).
17. Miledi, R. *J. Physiol. Lond.* **151**, 1–23 (1960).
18. Kuffler, S. W. & Yoshikami, D. *J. Physiol. Lond.* **224**, 703–730 (1975).
19. Fischbach, G. D. & Cohen, S. A. *Dev. Biol.* **31**, 147–162 (1973).
20. Nicolson, G. L. *Int. Rev. Cytol.* **39**, 89–190 (1974).
21. Nicolson, G. L. *Biochem. biophys. Acta* **457**, 57–108 (1976).
22. Brockes, J. P. & Hall, Z. W. *Biochemistry* **14**, 2100–2106 (1975).
23. Eylar, E. H., Madoff, M. A., Brody, O. V. & Oncley, J. L. *J. biol. Chem.* **237**, 1992–2000 (1962).
24. Poo, M-m. & Poo, W-j. H. *Science* (submitted).

letters to nature

Origin of pregalactic microwave background

THE photon density in the observed thermal (~ 2.7 K) microwave background is $n_{\gamma 0} \approx 400 \text{ cm}^{-3}$. This contrasts with the present mean baryon density in the Universe, n_b , which is in the range 10^{-5} – 10^{-6} cm^{-3} . In the standard 'hot big bang' cosmology, the ratio n_{γ}/n_b is essentially constant during the expansion, its value being

$$\mathcal{S} = n_{\gamma}/n_b \approx 3.6 \times 10^7 (\Omega h^2)^{-1}. \quad (1)$$

h denotes the Hubble constant H_0 in units of $100 \text{ km s}^{-1} \text{ Mpc}^{-1}$, and $\Omega = \frac{8}{3}\pi G\rho H_0^{-2}$ is the usual density parameter. $\mathcal{S}$ is an undetermined free parameter of the 'primordial fireball': the standard 'hot big bang' theory as yet gives no reason that $\mathcal{S}$ should be $\sim 10^8$ rather than (say) $\sim 10^4$ or $\sim 10^{12}$. Its observed value could be more readily understood in a different theory that attributes the radiative energy and entropy production to calculable processes in a specified epoch. There have been various proposals along these lines^{1–9}, but none has found as much favour as the standard 'hot big bang' cosmology. This is because production of the radiation at a redshift z requires a mass-energy conversion efficiency of the order

$$\rho_{\gamma}/\rho_b \approx 2.5 \times 10^{-5} (1+z)(\Omega h^2)^{-1} \quad (2)$$

This is implausibly high if z is large enough ($\gg 10^3$) to permit thermalisation by H and He plasma alone^{7–9}; on the other hand, if z corresponds to the epoch of galaxy formation (or later), adequate thermalisation at centimetre wavelengths requires implausibly efficient opacity^{3–5}. Here a possible non-primordial origin of the microwave background is outlined that seems less contrived than other such schemes.

For reasons outlined below, the dominant form of mass in the present Universe could be 'dead' remnants of massive or supermassive stars that formed when $t \leq 10^6$ yr. The active bright phase of these objects would last $\leq 10^7$ yr (and would be

over long before galaxies assembled as identifiable units) but could have generated enough radiation to satisfy equation (2) at $z \geq 100$. The uncondensed gas could, even at this early epoch, have been contaminated by the products of nucleosynthesis: it could contain heavy elements, molecules and dust, and would be dense and opaque enough to thermalise this radiation at a few hundred degrees, yielding a present-day microwave background with the characteristics observed.

To motivate a fuller astrophysical discussion of the radiation production and thermalisation, a general argument is outlined first which shows why this type of model naturally 'predicts' a value for $\mathcal{S}$ of order 10^8 .

Associated with any luminous gravitating objects where electron-scattering provides the dominant opacity is a critical luminosity, the 'Eddington luminosity' L_E , for which radiation pressure balances gravity. A characteristic timescale for conversion of rest mass into radiation is then¹⁰

$$\frac{Mc^2}{L_E} \doteq t_s = \frac{c\sigma_T}{4\pi Gm_p} \approx 4 \times 10^8 \text{ yr} \quad (3)$$

The lifetime of stars is related to this by

$$t = \epsilon \left(\frac{L}{L_E} \right)^{-1} t_s \quad (4)$$

ϵ being the efficiency of the energy source¹¹. For massive and supermassive stars $L \approx L_E$; and the expression (4) also gives the time taken for an accreting black hole of mass M to radiate ϵMc^2 .

Suppose that at some early time t_i some fraction F of the content of the Universe turns into self-gravitating objects which thereafter each radiate at a rate YL_E until they have radiated away a fraction ϵ of their mass-energy. This will take a time $(\epsilon Y^{-1})t_s$. At that time (assumed $\gg t_i$), the radiation energy density ρ_r will be $\frac{3}{8}\epsilon F\rho_b$, the factor $\frac{3}{8}$ arising from redshift effects. But ρ_b is related to t by $\rho_b = (6\pi Gt^2)^{-1}$ (for $(1+z) > \Omega^{-1}$). If the radiation were to be thermalised, its temperature T_{rad} would

be given by $aT_{\text{rad}}^4 = \rho_r c^2$. When the radiation constant a is written explicitly as $\pi^2 k^4 / 15 c^3 h^3$ this yields

$$\left(\frac{kT_{\text{rad}}}{m_p c^2}\right)^4 = \frac{18}{5} \left(\frac{Gm_p^2}{e^2}\right) \left(\frac{m_e}{m_p}\right)^4 \left(\frac{e^2}{hc}\right)^{-3} Y^2 \epsilon^{-1} F \quad (5)$$

and, as the photons have average energy $\sim 2.3kT$, $\mathcal{S} \approx [m_p c^2 / (2.3kT)] \rho_r / \rho_b$ which can be written

$$\mathcal{S} \approx \left(\frac{e^2}{Gm_p^2}\right)^{1/4} \left(\frac{m_p}{m_e}\right) \left(\frac{e^2}{hc}\right)^{3/4} Y^{-1/2} \epsilon^{5/4} F^{3/4} \quad (6)$$

This argument can be generalised to situations when Y and F are time-dependent. However, we will always find an expression for $\mathcal{S}$ which involves a fourth root of the ubiquitous 'large number' ($\sim 10^{40}$) multiplied by a set of model-dependent factors whose product is generally somewhat less than unity. It is thus not *ad hoc* that the class of non-primordial models discussed here yields background radiation with $\mathcal{S} \approx 10^8$; on the contrary, equation (6) displays an explicit reason that this should be expected.

Three specific astrophysical arguments pointing towards a phase of pregalactic star formation are: (1) data on their dimensions and spatial distribution suggest that galaxies have developed from density irregularities in the early Universe with a smooth spectrum of form $(\delta\rho/\rho) \propto M^{-\alpha}$ ($\frac{1}{2} \lesssim \alpha \lesssim \frac{3}{2}$) (refs 12–15). If this spectrum continued towards smaller scales, then gas clouds of $\lesssim 10^6 M_\odot$ would already have condensed out—collapsing (and possibly fragmenting) into star-like objects—at $t \lesssim 10^6$ yr. (This conclusion is insensitive to the Universe's thermal history at earlier times.)

(2) More than 80% of the material in the Universe seems to be 'dark mass' in galactic haloes and clusters of galaxies^{15–17}. This is consistent with the view^{18–20} that most matter was incorporated into one or more generations of O or B stars or supermassive objects which, after a 'brief' ($\sim 10^7$ yr) phase at luminosity $\sim L_E$, collapsed to form dark remnants. These remnants subsequently clustered gravitationally on progressively larger scales, and provided the potential wells in which the luminous content of galaxies eventually condensed¹⁵.

(3) Recent evidence that Fe has almost a solar abundance in the intergalactic medium^{21,22} suggests that pregalactic nucleosynthesis^{19,20} may have occurred.

In this hypothesis, several per cent of the rest mass could have been converted to radiation, through O or B stars with $L \approx L_E$ supernovae, recycling of debris through successive generations of exploding stars or supermassive objects, accretion of dense surrounding gas on to compact remnants, and so on. Thus, even when $t \lesssim 10^7$ yr ($z \gtrsim 100$), ρ_r/ρ_b can be lifted to the value of equation (2), so that—given adequate thermalisation—it can provide the whole microwave background.

Nucleosynthesis could have supplied heavy elements by $t \sim 10^7$ yr (even if they are not produced 'primordially' in some non-standard 'big bang'). At $z \gtrsim 100$ the mean densities and probable composition throughout the entire Universe resemble interstellar clouds in our present-day Galaxy: we may expect molecules, and carbon or silicate grains. The radiation densities and path lengths are, however, orders of magnitude larger, and if the radiation were thermalised its temperature would be several hundred degrees.

If thermalisation processes can produce a 2.7 K background that is thermal at a present wavelength λ_0 (and we know the background is close to a black body spectrum at least for $\lambda_0 \lesssim 30$ cm) then the pregalactic medium must have been opaque at wavelengths $\lambda_0(1+z)^{-1}$ where $z \gtrsim 100$ and $T \gtrsim 270$ K.

Three contributors to the opacity are dust, molecules and the ionised component of the medium.

First, dust: the effective absorption cross-section of a grain of radius r can be written $Q(\lambda)\sigma_G$, σ_G being the geometrical cross-section. When $\lambda \gg r$, generally $Q(\lambda) \ll 1$. In fact, for spherical grains one expects $Q(\lambda) \lesssim Q_{\text{max}}(\lambda) = 4\pi(r/\lambda)$. This means that if $Q(\lambda) \approx Q_{\text{max}}(\lambda)$ the column density of material needed to yield optical depth unity at a wavelength $\lambda (\gg r)$ is

proportional to λ but independent of r : for grains with typical density $\sim 2 \text{ g cm}^{-3}$ it is $\sim 2/3(\lambda/1 \text{ cm}) \text{ g cm}^{-2}$. The characteristic column density of the Universe at redshift z is $\sim 0.14(1+z)^{3/2} \Omega^{1/2} h \text{ g cm}^{-2}$. To provide optical depth $\gtrsim 1$ at redshift z for radiation whose present wavelength is λ_0 the fraction of material in the form of grains must satisfy

$$\frac{\rho_{\text{grain}}}{\rho} \gtrsim 5 \left(\frac{\lambda_0}{1 \text{ cm}}\right) (1+z)^{-5/2} \frac{Q_{\text{max}}(\lambda_0(1+z)^{-1})}{Q(\lambda_0(1+z)^{-1})} \Omega^{-1/2} h^{-1} \quad (7)$$

Q is higher at shorter wavelengths, so whenever equation (7) holds at centimetre wavelength any radiation emitted in the optical or ultraviolet (UV) band will be quickly transformed (perhaps by a sequence of absorptions and re-emissions) into softer radiation and adequately thermalised within the expansion timescale for the Universe.

Second, molecules: if some heavy elements are combined into simple molecules (H_2O , CO , CS , and so on) then the lines corresponding to the rotational transitions will be smeared into overlapping continua by the Hubble expansion. If the levels are in thermal equilibrium (and so populated in accordance with statistical weights when separated by energies $< kT$) the condition for a given molecule, with fractional abundance n_{mol}/n and dipole moment $\mu \text{ Db}$, to provide $\tau > 1$ at $\lambda_0(1+z)^{-1}$ is²⁴

$$n_{\text{mol}}/n \gtrsim 5 \times 10^{-4} \mu^{-2} (1+z)^{-3/2} (\lambda_0/1 \text{ cm}) \quad (8)$$

Third, free-free absorption by the ionised component: the primary sources of the radiant energy may supply UV photons capable of ionising the pregalactic medium. If there is sufficient dust to contribute to thermalisation, it will have a very large optical depth to these photons (see ref. 7). The grains are therefore very efficient at converting the energy of one 'hard' photon into many softer ones. Even though this means that most of the radiation energy density will be in the IR rather than the UV, there may, nevertheless, be enough photo-ionisations to maintain n_e/n at a significant fraction. The high density of soft photons has the interesting consequence that the Compton cooling time $t_{\text{comp}} \approx 3 \times 10^{11} (\tau_{\text{rad}}/270 \text{ K})^{-4} \text{ s}$ may be less than the recombination time $t_{\text{rec}} \approx 2 \times 10^{11} n_e^{-1} (T/270 \text{ K})^{1/2} \text{ s}$. In these conditions, the electrons may have $T_e \approx T_{\text{rad}} (< 10^4 \text{ K})$ even when n_e/n is high. Free-free absorption, important only at long wavelengths, will then enhance the thermalisation rather than distorting the spectrum²⁵. The free-free optical depth exceeds unity for

$$\lambda_0 \gtrsim 2 \times 10^2 h^{-3/2} (1+z)^{-1/2} \Omega^{-3/2} \left(\frac{n_e}{n}\right)^{-1} \left(\frac{\bar{n}_e^2}{n_e^2}\right)^{1/2} \times \left(\frac{T_e}{2.7(1+z)}\right)^{3/4} \text{ cm} \quad (9)$$

but the value of n_e is too model-dependent to permit any firm estimate of how important this process is.

One example will show that thermalisation is no great problem for this general scheme. Suppose that $\Omega = h = 1$, and that 0.5% of the mass energy has been converted into radiation when $t \approx 3 \times 10^6 \text{ yr}$ ($z \approx 200$). This satisfies equation (2), and corresponds to a radiation rate of order L_E , assuming that most of the mass is involved in the pregalactic phase of star formation. Thermalisation at $\lambda_0 < 30 \text{ cm}$ would be guaranteed even if only $2.5 \times 10^{-4} (Q_{\text{max}}/Q)$ of the material were incorporated in grains. The actual value of Q depends on grain properties, but small grains of amorphous silicates may have $Q \approx Q_{\text{max}}$ at all wavelengths up to $\sim 1.5 \text{ mm}$ (refs 26, 27) (the wavelength at $z = 200$ corresponding to 30 cm at the present epoch). There is thus no need to invoke grains of exotic shapes or constitutions^{3,4}. Equation (8) suggests that molecules may be important: one requires only 3×10^{-6} (by number) in the form of some molecule with $\mu \approx 1$. The considerations leading to equation (9) suggest that free electrons in the pregalactic medium would be cooled by Compton drag to kinetic temperatures of a few hundred degrees: this provides a favourable environment for molecule formation²⁸.

The detailed physics of the thermalisation (and the epoch when it occurs) are quite uncertain: all that can be argued here is

that the general scheme is plausible in order-of-magnitude terms. The following types of observation are, however, relevant. (1) Background spectrum: although the spectrum should be essentially thermal, deviations from the Planck law might arise if energy were injected at epochs when thermalisation was no longer fully effective at all wavelengths. Also, if some volatile contributor to the opacity with a spectral feature at some wavelength λ_1 were to condense out at a well-defined temperature T_1 , then the background spectrum might now display a feature at wavelength $\lambda_1(T_1/2.7\text{ K})$. (2) Small-scale isotropy: the effective source of the microwave photons reaching us would lie at $z \geq 100$, when they were emitted by a grain or molecule, or scattered by a free electron. The expected amplitude of the temperature anisotropies due to inhomogeneities would be similar to those predicted by the 'hot big bang' picture. An interesting difference in the present scheme, however, stems from the fact that various wavelength-dependent opacities (and not just electron scattering) are important: this means that the angular fluctuations observed at different wavelengths need not be correlated, as the location of the 'cosmic photosphere' depends on wavelength. (3) Galaxy formation and 'missing mass': any observations bearing on galaxy formation, the value of Ω , and the nature of the 'dark' mass in galactic haloes and clusters are relevant to the ideas discussed here.

In conclusion, there are astrophysical reasons for expecting a large radiative output from pregalactic events when $t \leq 10^7$ yr. This radiation could be thermalised, and would now be comparable in temperature to the observed microwave background. It is thus tempting to suppose that essentially all this background originated in this way rather than being a vestige of a hot 'primordial fireball'. The general argument leading to equation (6) then allows us to understand why the parameter $\mathcal{P}$ —whose value is unexplained by the 'hot big bang' theory—is $\sim 10^8$. The arguments pertaining to helium and deuterium are thus our only real clues to the entropy and dynamics at much earlier eras; and the options of non-zero lepton numbers or non-Friedmannian dynamics complicate this issue (as do the uncertainties about nucleosynthesis in the pregalactic objects themselves). The very early Universe thus seems a topic for conjecture rather than consensus while such obscurity veils the first 3 Myr.

I thank Michael Rowan-Robinson, Bernard Carr, Bernard Jones and Simon White for helpful comments.

MARTIN J. REES

*Institute of Astronomy,
Madingley Road,
Cambridge, UK*

Received 1 June; accepted 24 July 1978.

1. Layzer, D. *Astrophysics and General Relativity* 2, 155–232 (Gordon and Breach, New York, 1969).
2. Kaufman, M. *Astrophys. J.* **160**, 459 (1970).
3. Layzer, D. & Hively, R. M. *Astrophys. J.* **179**, 361 (1973).
4. Narlikar, J. V., Edmunds, M. G. & Wickramasinghe, D. in *Infrared Astronomy* (ed. Rowan-Robinson, M.) (Pergamon, Oxford, 1975).
5. Aliven, H. & Mendis, A. *Nature* **266**, 699 (1977).
6. Pollaine, S. *Nature* **271**, 426 (1978).
7. Rees, M. J. *Phys. Rev. Lett.* **28**, 1669 (1972).
8. Carr, B. J. & Rees, M. J. *Astr. Astrophys.* **61**, 705 (1977).
9. Carr, B. J. *Astr. Astrophys.* **60**, 13 (1977).
10. Salpeter, E. E. *Astrophys. J.* **140**, 796 (1964).
11. Dicke, R. H. *Nature* **192**, 440 (1962).
12. Peebles, P. J. E. *Astrophys. J. Lett.* **189**, L51 (1974).
13. Peebles, P. J. E. & Dicke, R. H. *Astrophys. J.* **154**, 891 (1968).
14. Gott, J. R. & Rees, M. J. *Astr. Astrophys.* **45**, 365 (1975).
15. White, S. D. M. & Rees, M. J. *Mon. Not. R. astr. Soc.* **183**, 341 (1978).
16. Einasto, J., Kaasik, A. & Saar, E. *Nature* **250**, 309 (1974).
17. Ostriker, J. P., Peebles, P. J. E. & Yahil, A. *Astrophys. J. Lett.* **193**, L1 (1974).
18. Truran, J. W. & Cameron, A. G. W. *Astrophys. Space Sci.* **14**, 179 (1971).
19. Hartquist, T. W. & Cameron, A. G. W. *Astrophys. Space Sci.* **48**, 145 (1977).
20. Rees, M. J. *Physica Scripta* **17**, 371 (1978). 1. Mitchell, R. J., Culhane, J. L., Davison, P. J. N. & Ives, J. C. *Mon. Not. R. astr. Soc.* **176**, 29 (1976a).
22. Mushotsky, R. F., Serlinitos, P. J., Smith, B. W., Boldt, E. A. & Holt, S. S. *Astrophys. J.* (in the press).
23. Purcell, E. M. *Astrophys. J.* **158**, 433 (1969).
24. Goldreich, P. & Kwan, J. *Astrophys. J.* **189**, 441 (1974). 5. Sunyaev, R. A. & Zeldovich, Y. B. *Astrophys. Space Sci.* **7**, 20 (1970).
25. Sunyaev, R. A. & Zeldovich, Y. B. *Astrophys. Space Sci.* **7**, 20 (1970).
26. Day, K. L. *Astrophys. J.* **210**, f4 (1976).
27. Aannestad, P. A. *ophys. J.* **220**, 538.
28. Herbst, E. & Klempner, W. *Astrophys. J.* **185**, 505.

Venus' rotation and atmospheric tides

VENUS rotates with a period of 243.00 ± 0.04 d retrograde¹. The obliquity, or angle between the spin vector and orbit vector is $178 \pm 1^\circ$ (ref. 1). Both the long period and the near 180° obliquity suggest that the spin has evolved under the influence of tidal torques. Tides raised by the Sun in the body of Venus would de-spin the planet in $\sim 10^8$ yr if no other torques were acting². The final state would be one of synchronous rotation (one side always facing the Sun). So either we are living during the final stages of Venus' tidal evolution, an unlikely circumstance, or else Venus has already reached a stable equilibrium in which other influences balance the solar body tide. It is possible that the Earth has 'captured' Venus into a resonance in which the spin period is 243.165 d. However, for this resonance to be a stable equilibrium, either Venus must have a gravitational field that is ~ 10 times 'rougher' than the Earth's², a possibility that is largely ruled out by direct observation³, or a third torque must balance the body torque due to the Sun. Tides in the atmosphere driven by periodic solar heating could supply the necessary third torque^{4–6}, but no quantitative theory has previously been published. Such a theory⁷ is presented here, in which we argue that the current rotation is a stable balance between atmospheric and solar body tides.

Because the atmosphere is thin compared with the radius, tidal torques on the atmosphere depend only on lateral variations of atmospheric column density, which is proportional to surface pressure. Although less energy is absorbed at the surface than in the clouds (100 W m^{-2} compared with 500 W m^{-2} at normal incidence)⁸, heating at the ground is the dominant source of surface pressure variations. This follows from the 25-fold decrease in solar diurnal period, from 117 d at the surface to 4 d in the clouds⁸, and the fact that surface pressure oscillations vary roughly as $(\text{period})^2 \times (\text{pressure})^{-1/2}$ at the level of absorption.

It is not difficult to show that the power absorbed by the ground is quickly exchanged with the lower atmosphere. A simple model, representative of more distributed forcing, treats this heating as being confined to a thin layer above the surface. Because of the rigid lower boundary condition, such heating does not excite significant tidal oscillations in the atmosphere above, and fluid parcels within the heated layer remain approximately at constant pressure. The rate of change of column density is therefore

$$\frac{d}{dt} \int_0^\infty \rho dz = - \int_0^{z_0} \frac{\rho}{T} \frac{dT}{dt} dz = - \frac{F}{c_p T_0} \quad (1)$$

where z_0 is the thickness of the heated layer, t is time, ρ is atmospheric density, T_0 is surface temperature, c_p is specific heat at constant pressure, and F is the heat flux (in W m^{-2}) absorbed by the layer. To simulate the finite residence time of a heat pulse in the lower atmosphere we replace d/dt by $d/dt + 1/\tau$ where τ is a thermal time constant associated with boundary layer processes and is in general a function of the forcing frequency.

Equation (1) is solved for the column density as a function of planetocentric latitude, longitude, and time, assuming the lower atmosphere rotates with the solid body. The heat flux F is expressed as a function of these variables by assuming $F = 0$ on the night side and $F = F_0 \cos \zeta$ on the day side, where ζ is the solar zenith angle. The column density is then multiplied by the gradient of the tidal potential, crossed with the planetary radius vector and integrated over the spherical surface to give the net torque on the atmosphere. In a steady state, this torque must be transmitted to the crust. The torque may be resolved into a component parallel to the spin axis, which affects the magnitude ω of the spin angular velocity, and a perpendicular component which affects the obliquity. Defining the angle β as

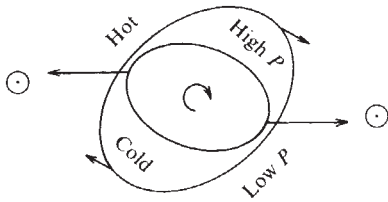


Fig. 1 Balance of torques on Venus. As only the semidiurnal components of the tides produce a torque, an 'image Sun' is shown diametrically opposite the true Sun, to make the picture symmetrical and to aid visualisation. The Sun is held fixed in this sketch so that Venus is rotating clockwise as seen from north of its orbit plane. The inner oval represents the tidally distorted figure of Venus, while the outer oval represents a surface of constant atmospheric pressure; the outer oval may also be regarded as the surface of an ocean of equivalent mass. Solar radiation raises the air temperature in the afternoon and causes mass to flow to the colder morning region. The high-pressure areas as well as the tidal bulges in the body of Venus are fixed with respect to the Sun (but not with respect to the surface). The gravitational attraction of the Sun, represented by arrows in the diagram, thus exerts contrary torques on the atmosphere and body of Venus.

the supplement of the obliquity, for $\beta \ll 1$ (as at present) one obtains

$$\frac{d\omega}{dt} = \left[\frac{3\pi n^2 r_0^4 F_0}{8c_p T_0 C} \right] \delta(2\omega + 2n) \quad (2)$$

$$\frac{1}{\beta} \frac{d\beta}{dt} = \left[\frac{3\pi n^2 r_0^4 F_0}{8c_p T_0 C \omega} \right] \frac{1}{2} [-\delta(2\omega + 2n) - \delta(\omega + 2n) + \delta(\omega)] \quad (3)$$

where

$$\delta(\sigma) = \sigma(\sigma^2 + \tau^{-2})^{-1} \quad (4)$$

Here n is the orbital angular speed, r_0 is the planetary radius, and C is the planet's greatest moment of inertia.

Equations (1)–(3) have a simple physical interpretation. Heating at the ground causes temperature to rise and column density to fall (equation (1)). Hence there is a mass concentration at the coldest time just after sunrise and a mass deficit at the hottest time just before sunset. The Sun's gravity acting on the sunrise side leads to a torque that tends to speed up the atmosphere (Fig. 1). The relevant frequency $2\omega + 2n$ is that of the solar semi-diurnal tide (equation (2)). The obliquity is affected first by the component of the semi-diurnal tidal torque that acts perpendicular to the spin axis (frequency $2\omega + 2n$), and second by the tides at frequencies ω and $\omega + 2n$ that result from the annual migration (at frequency n) of the sub-solar point across the equator, modulated (at frequency $\omega + n$) by the motion of the Sun relative to the surface (equation (3)).

The gravitationally induced tide that acts on the solid planet leads to expressions of form similar to those above²:

$$\frac{d\omega}{dt} = \left[\frac{3k_2 GM_{\odot}^2 r_0^5}{2a^6 C} \right] [-\varepsilon(2\omega + 2n)] \quad (5)$$

$$\frac{1}{\beta} \frac{d\beta}{dt} = \left[\frac{3k_2 GM_{\odot}^2 r_0^5}{2a^6 C \omega} \right] \frac{1}{2} [\varepsilon(2\omega + 2n) + \varepsilon(\omega + 2n) - \varepsilon(\omega)] \quad (6)$$

(Symbols are defined in ref. 2.) The $\varepsilon(\sigma) \approx Q^{-1}$ are phase lags with the same sign as σ that depend on dissipation in the solid planet. $Q(\sigma)$ is the quality factor for oscillations at frequency σ . Values of Q for Venus are unknown, but analogy with the Earth suggests $30 \leq Q \leq 100$ (ref. 9).

According to equations (2) and (5), atmospheric tides tend to increase ω away from synchronous rotation, while solar body tides tend to drive the spin toward synchronism. For $\tau \rightarrow \infty$ (long thermal response time), balance at the current diurnal period of 117 d requires $Q \approx 30$, as shown in Fig. 2. For smaller τ , balance at the current period requires larger Q . Equations (3) and (6) indicate that the current state is stable to obliquity

perturbations provided τ is less than about 2 weeks. The Earth resonance is also stable provided atmospheric and solar body tides balance at a period sufficiently close to the resonant period.

We have shown in which conditions the current rotation might be a stable equilibrium. As shown in Fig. 2, there is also a prograde equilibrium state with a diurnal period of 117 d (spin period ≈ 77 d). The stability of this state may be analysed by substituting $\pi - \beta$ for β and $-n$ for n in equations (2)–(6). Our analysis shows that when one of the two equilibrium states in Fig. 2 is stable to obliquity perturbations, the other is unstable.

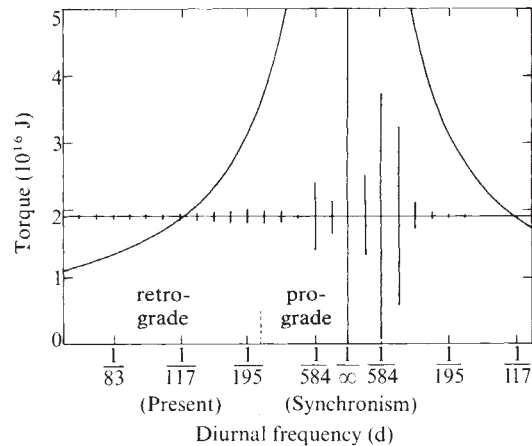


Fig. 2 Tidal torques plotted against the diurnal frequency $(\omega + n)(2\pi)^{-1}$ for spin vector parallel to the orbit vector. Atmospheric torques (curved lines) tend to drive the spin away from synchronism, and are shown for values of the thermal time constant $\tau \rightarrow \infty$. The solar tidal torque on the solid body (horizontal line) tends to drive the spin toward synchronism, and is shown for a value of the quality factor $Q = \text{constant} \approx 30$. This choice of Q ensures that atmospheric and solid body torques together drive the spin toward the current diurnal period of 117 d. Torques due to Earth resonances (vertical lines) tend to hold the spin period at the resonant frequencies, and are shown for a value of $(B - A)C^{-1} = 2.2 \times 10^{-5}$, similar to that for the Earth. Here A, B, C are the three principal moments of inertia of Venus, with $A < B < C$. For this amount of gravitational roughness, the Earth torques are too small by an order of magnitude to overcome the solar solid body torques, unless the solar torques and atmospheric torques balance near a resonant frequency. A recent analysis¹ of radar data suggests that the spin period is very close but not equal to that required for resonance with the Earth.

However, in general, other equilibria (both stable and unstable) can occur with obliquities between 0° and 180° . A more complete theory to be presented later indicates that Venus probably originated with a retrograde rotation in order to have evolved to the current retrograde state.

This research was supported by NASA.

ANDREW P. INGERSOLL

ANTHONY R. DOBROVOLSKIS

Division of Geological and Planetary Sciences,
California Institute of Technology,
Pasadena, California 91125

Received 3 April; accepted 29 June 1978.

- Shapiro, I., Campbell, D. & DeCampi, W. *Astrophys. J. Lett.* (submitted).
- Goldreich, P. & Peale, S. *Astr. J.* **72**, 662–668 (1967); **75**, 273–284 (1970).
- Howard, H. T. *et al. Science* **183**, 1297–1300 (1974).
- Gold, T. & Soter, S. *Icarus* **11**, 356–366 (1969).
- Hinch, E. J. Lecture notes from the Woods Hole Oceanographic Institution Summer Study Program in Geophysical Fluid Dynamics (1970).
- Kundt, W. *Astr. Astrophys.* **60**, 85–91 (1977).
- Ingersoll, A. & Dobrovolskis, A. *Bull. Am. astr. Ass.* **9**, 467 (1977).
- Keldysh, M. V. *Icarus* **30**, 605–625 (1977).
- Munk, W. & MacDonald, G. *The Rotation of the Earth* (Cambridge University Press, 1975).

Optical detection of local interstellar particles during an out-of-ecliptic mission

THE contribution to the zodiacal light (ZL) by interstellar grains has been estimated¹⁻³, but, as suggested here, new observations and analyses of the ZL and of the local interstellar medium provide a firmer basis for placing limits on the brightness of the interstellar (IS) relative to the interplanetary (IP) component. The proposed NASA/ESA 'out-of-ecliptic' (or 'solar polar') mission⁴ provides the best and perhaps the only observational method to distinguish between these two components.

To estimate the interstellar dust density, we make the usual assumption that the number density of the interstellar particles (n_{IS}) is related to the local hydrogen number density in the ratio of 10^{-12} (see ref. 5). The local hydrogen density within a few parsecs of the Sun has been observed to be $\sim 0.1 \text{ cm}^{-3}$ (ref. 6) which suggests that we are currently passing through an intercloud region of the galaxy. We infer from this gas density that the IS dust content is about $n_{IS} = 10^{-13} \text{ cm}^{-3}$.

The effect of the Solar System on the ambient IS complex of gas and dust is unknown. Although several mechanisms have been proposed either for capture^{7,8} or the sweeping out of interstellar grains, none of the results are conclusive because of uncertainties in the basic assumptions. The most efficient mechanism for deflecting interstellar particles (if charged) is probably the Lorentz force exerted by the magnetic field associated with the solar wind⁹. Levy and Jokipii¹⁰ estimate the ratio of the Lorentz to the gravitational force (F_L/F_G) at 1 AU to be 10, but their use of an IS particle size of $0.02 \mu\text{m}$ is too small by at least a factor of five for the classical size particles¹¹. If a realistic particle size of characteristic radius $0.12 \mu\text{m}$ is used, a value of F_L/F_G of the order of 0.4 (rather than >1) is derived, and sweeping out of the IS particles should not occur. Given the absence of detailed dust orbit calculations as well as the uncertainties in both the initial grain charge properties and in the possibility of the grains discharging in the IP medium, we base our calculations on the assumption that the grains are neither accreted nor swept out, but that the IS grains in the Solar System are uniformly distributed with a density characteristic of the local IS medium.

We estimate the monochromatic IS contribution to the ZL by scattering of solar radiation from an arbitrary distribution of interstellar particles from the brightness integral¹²

$$Z_{IS} = \int_{LOS} f_{\odot} \left(\frac{R_{\odot}}{R} \right)^2 n_{IS}(\theta) dl \quad (1a)$$

which becomes, for a uniform IS density,

$$Z_{IS} = n_{IS} f_{\odot} R_{\odot}^2 \int_{LOS} \frac{\sigma_{IS}(\theta)}{R^2} dl \quad (1b)$$

where LOS is line-of-sight, $f_{\odot}(R_{\odot}/R)^2$ is the solar flux on a typical volume element at a heliocentric distance R , θ is the scattering angle, $\sigma_{IS}(\theta)$ is the differential scattering cross section of the IS dust, and dl is measured along the LOS. The integral in equation (1b) is calculated from

$$\int_{LOS} \frac{\sigma_{IS}(\theta)}{R^2} dl = \frac{1}{b} \int_{\epsilon}^{\pi} \frac{F(\theta)}{k^2} d\theta \quad (2)$$

where b is the closest solar approach of the LOS, ϵ is the elongation angle (see Fig. 1), and k is the wave number ($2\pi/\lambda$) of the incident radiation. For the scattering function $F(\theta)$, we use a dielectric dust model¹³. At a wavelength of $4,400 \text{ \AA}$, the conversion of cgs radiance units to the unit $S_{10}(V)$ (using the observational parameters recommended by Sparrow and Weinberg¹⁴) is

$$S_{10}(V) = 1.2 \times 10^{-9} \text{ erg cm}^{-2} \text{ s}^{-1} \text{ sr}^{-1} \text{ \AA}^{-1} \quad (3)$$

where $S_{10}(V)$ is defined as the radiance produced by a 10th visual magnitude solar (G2V) star per square degree.

According to the model shown in Fig. 1, at the Earth ($b = 1 \text{ AU}$ and for $\epsilon = 90^\circ$), the IS contribution to the ZL is only about $1 S_{10}(V)$. This value is not only below the limit of detectability of $\sim 2-3 S_{10}(V)$ (ref. 15); it is also totally masked by the $200 S_{10}(V)$ contributed by the IP particles. On the other hand, viewing sunwards from a space craft (located at $R = 1 \text{ AU}$) along an LOS parallel to the ecliptic plane and $1/2 \text{ AU}$ above it (out-of-the-ecliptic), the IS brightness estimated from equation (2) for the then appropriate value of $\epsilon = 27^\circ$ is $\sim 28 S_{10}(V)$ which is well above the limit of detectability. Furthermore, this IS contribution may be distinguishable from the IP component.

An observational upper limit to the IS brightness component is established by noting that the ZL is close to solar colour from about $2,200 \text{ \AA}$ to $2.4 \mu\text{m}$ (ref. 16 and refs therein). Since the IS particle scattering is colour (wavelength) dependent in this range, this may be used to establish an upper limit to the IS component. Our calculations are confined to the wavelength spread of the UBV system where more data are available for separating the ZL and the background starlight. The ZL in this wavelength region is known to be solar colour to within about 5%.

Assume that the ZL consists of two components, the IP part which is solar colour and the IS component whose wavelength dependence follows the IS dust model previously cited. Thus if we consider the two wavelengths $5,400 \text{ \AA}$ and $3,600 \text{ \AA}$, the IP component has a spectral ratio $I_{IP}(5,400)/I_{IP}(3,600) = 1(I_{IP} \neq I_{IP}(\lambda))$; I is normalised to solar colour). According to Greenberg and Hanner¹³, the IS component has a spectral ratio $I_{IS}(5,400)/I_{IS}(3,600) = 0.77$ at $\epsilon = 90^\circ$.

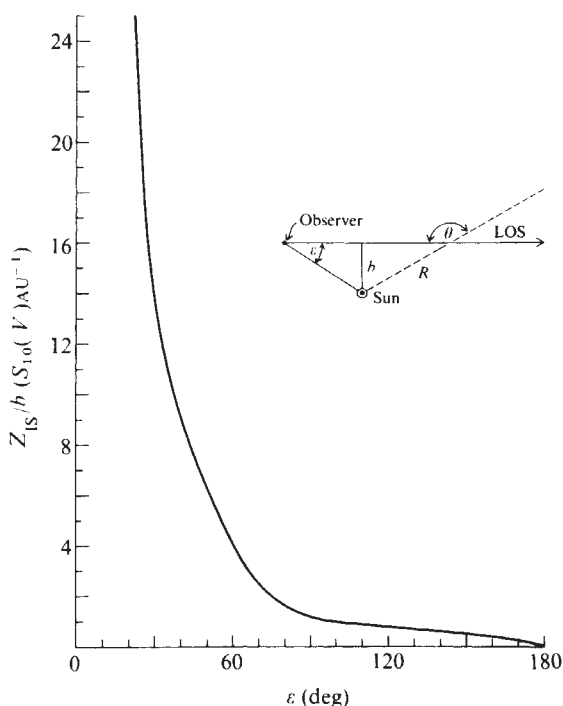
The 5% colour variation constraint then implies

$$\frac{I_{IP} + I_{IS}(3,600)}{I_{IP} + I_{IS}(5,400)} = \frac{1 + I_{IS}(3,600)/I_{IP}}{1 + 0.77 \times I_{IS}(3,600)/I_{IP}} \leq 1.05 \quad (4a)$$

$$\text{so that} \quad \frac{I_{IS}(3,600)}{I_{IP}} \leq 0.25 \quad \text{or} \quad \frac{I_{IS}}{I_{ZL}} \leq 0.20 \quad (4b)$$

This means that no more than about 20% of the total zodiacal light as viewed in the ecliptic and at 90° elongation can be produced by scattering off of interstellar type particles.

Fig. 1 Resulting ZL brightness in $S_{10}(V)$ units per unit (AU) distance of closest solar approach along the line of sight from $\theta = \epsilon$ to $\theta = \pi$. Inset shows appropriate geometry.



As knowledge of the size and spatial distribution of the IP particles out of the ecliptic is lacking, we derive the maximum amount of IP dust $\frac{1}{2}$ AU off the ecliptic that would still allow the IS component to be distinguished. Let $f \cdot n_{IP}$ be this amount, where n_{IP} is the dust density in the neighbourhood of the Earth and f is the fraction to be determined. As a reference point, the neighbourhood of the Earth has a volumetric scattering cross section of $n_{IP} \sigma_{IP}(90^\circ) = 7.8 \times 10^{-23} \text{ cm}^{-1} \text{ sr}^{-1}$ (ref. 17). For $\theta = 45^\circ$, the scattering cross section $\sigma_{IP}(45^\circ)$ is about $1.77 \sigma_{IP}(90^\circ)$ (ref. 18).

Our minimum distinguishability criterion is that the ratio of the interstellar to interplanetary contributions be greater than 25%. For the case treated above for the interstellar contribution at $\frac{1}{2}$ AU above the ecliptic, then

$$\frac{Z_{IS}}{Z_{IP}} \approx \frac{n_{IS} \sigma_{IS}(45^\circ)}{n_{IP} \sigma_{IP}(45^\circ)} = \frac{7.5 \times 10^{-11} \times 10^{-13}}{f \times 1.77 \times 7.8 \times 10^{-23}} \geq 0.25$$

from which we obtain $f \leq 0.20$.

This value of f is close to but smaller than those derived from a series of models by Giese¹⁹ which produced results consistent with those inferred from observations from the Earth. Most of these models predict the dust density (assumed to be purely interplanetary) to decrease to 20% between heights of 0.5 and 1 AU above (or below) the ecliptic.

These rough calculations suggest there is a real possibility of optically detecting a local IS dust component in the Solar System. Although our criterion of distinguishability is based on colour arguments alone, both polarisation and scattering angle dependence provide additional tools to distinguish between the larger (IP) and smaller (IS) particles. The chance of isolating the IS component increases with decreasing elongation angle. Observations should be made along a line of sight which is parallel to the ecliptic and is as close to the Sun as possible. Fortunately, such observations can also be inverted; not only can the total brightness (Z) be obtained, but when the spacecraft is near maximum height the value $n\sigma$ can be directly determined from the measurements²⁰.

Assuming that a positive identification of interstellar dust will be made, we are on the threshold of observing physical parameters of these particles which could previously only be inferred from observations of stellar extinction and scattering in reflection nebula, which are normally hundreds of parsecs from our local neighbourhood. For example, given a slightly more optimistic estimate (by perhaps a factor of 2) of the local interstellar dust density, we might be able to observe how they are affected by interaction within the Solar System.

D. S. acknowledges the support of NASA and the US Air Force Office of Scientific Research. We thank Dr R. H. Giese and Dr J. L. Weinberg for useful suggestions.

J. MAYO GREENBERG

Laboratory Astrophysics,
Leiden University,
The Netherlands

D. W. SCHUERMAN

Space Astronomy Laboratory,
State University of New York at Albany.

Received 19 June; accepted 14 July 1978.

- Öpik, E. J. *Proc. R. Irish Acad.* **54A**, 165 (1951).
- Greenberg, J. M. *Space Res.* **9**, 445 (1969).
- Lillie, C. F. *The Scientific Results from OAO2*, (ed. Code, A. O.) 95 (NASA SP-311, 1972).
- Fisk, L. A. & Axford, W. I. (eds) *Proc. Symp. Study of the Sun and Interplanetary Medium in Three Dimensions* (NASA Publ. X-660-75-53, 1976).
- Greenberg, J. M. in *Cosmic Dust*, (ed. McDonnell, J. A. M.) chap. 4, 187 (Wiley, London, 1978).
- McIntock, W., Henry, R. C., Moos, H. W. & Linsky, J. L. *Astrophys. J. Lett.* **204**, L103 (1976).
- Lyttleton, R. A. *The Comets and Their Origin* (Cambridge University Press, 1953).
- Gary, G. A. *The Contribution of Interstellar Particles to the Interplanetary Dust Complex*, (NASA Tech. Mem., NAS TM X-73377, 1977).
- Greenberg, J. M. *Proc. IAU Colloq.* No. 13 (1973).
- Levy, E. H. & Jokipii, J. R. *Nature* **264**, 424 (1976).
- Greenberg, J. M. & Hong, S. S. *Eislab Symp.* No. 8, Fascati (ed. Moorwood, A. F. M.) (ESRO SP 105, 153, 1974).
- Leinert, C. *Space Sci. Rev.* **18**, 281 (1975).
- Greenberg, J. M. & Hanner, M. S. *Astrophys. J.* **161**, 947 (1970).
- Sparrow, J. G. & Weinberg, J. L. *Lecture Notes in Physics*, No. 48, 41, (Springer, Heidelberg, 1976).

- Schuerman, D. W., Weinberg, J. L. & Beeson, D. E. *Bull. Am. Astr. Soc.* **9**, 313 (1977).
- Weinberg, J. L. & Sparrow, J. G. in *Cosmic Dust* (ed. McDonnell, J. A. M.) (Wiley, London, 1978).
- Dumont, R. *Planet. Space Sci.* **21**, 2149 (1973).
- Giese, R. H., Weiss, K., Zerull, R. H. & Ono, T. *Astr. Astrophys.* (in the press).
- Giese, R. H. *Eldo-Cecles/Esro-Cers scient. tech. Rev.* **7**, 43 (1975).
- Schuerman, D. W. *Bull. Am. Astr. Soc.* **9**, 564 (1978).

Extent of hot intergalactic gas in the cluster Abell 2218

A SMALL cooling of the microwave background towards the centre of the rich, distant cluster of galaxies Abell 2218 has been detected¹⁻³. This decrement has been attributed to the inverse Compton scattering of photons of the background radiation by electrons in a hot intracluster plasma, as predicted by Sunyaev and Zel'dovich⁴. We report here new observations in several different directions towards the cluster which confirm the earlier results and provide an estimate of the extent of the gas.

The new measurements were made during the period February 1977 to May 1978 at the SRC Chibolton Observatory, and consist of 199 h of observation at points north and south of the Abell cluster centre⁵ and 54 h further at that position. As in previous measurements^{1,3}, these observations were made at prime focus of the 25-m telescope operated at 10.6 GHz, using a twin-beam system with a beam-throw of 15 arc min in azimuth and HPBW of 4.5 arc min. The receiver incorporated a parametric amplifier with noise temperature about 180 K and bandwidth 250 MHz. The experimental method and data-reduction have been described in detail elsewhere³.

Table 1 shows the coordinates of the positions observed and the temperature decrements measured there. No corrections for the presence of nearby radio sources are needed⁶. The present results are consistent with earlier measurements near dec $66^\circ 20'$ —as discussed by Birkinshaw, Gull and Northover³. The measurement at dec $66^\circ 18'$ was made to check on the strong detections at $66^\circ 16'$ and $66^\circ 20'$ and to furnish another point close to the cluster centre. Its presence leads to an appreciable

Fig. 1 The galaxy strip count, N , in arc min bins and the observed temperature decrement at each point, ΔT , plotted as functions of declination, δ (1950.0 coordinates). The solid line is the ΔT profile of the best-fitting $\gamma = 5/3$, $\xi = -0.15$ atmosphere convolved with the beamshape of the Chibolton telescope. The linear scale, D , at the distance of the cluster is also marked.

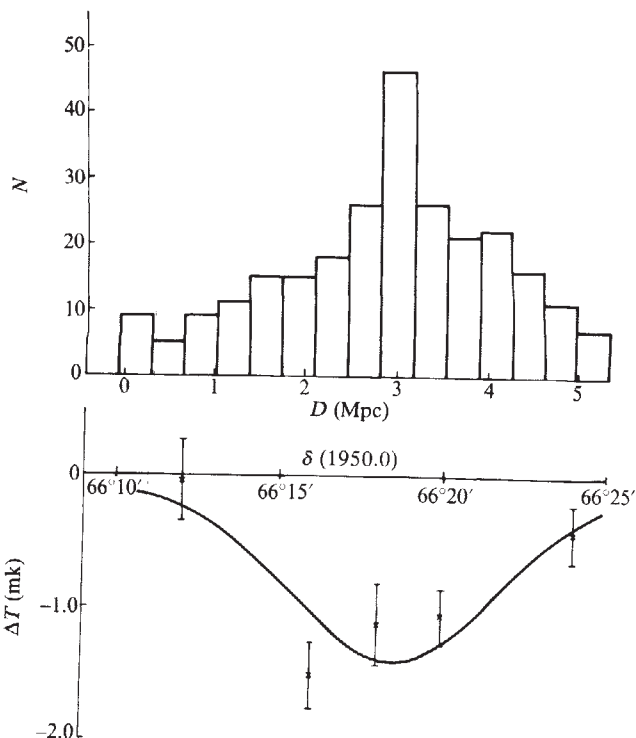


Table 1 Measured temperatures towards the observing positions in Abell 2218

Observing coordinates (1950.0)	Distance from cluster centre (arc min)	Time of observation (h)	Result (mK)	$\sqrt{\chi^2} - \sqrt{n-1}$
RA Dec				
16h 35min 42s 66° 12'	6.5	38	-0.03 ± 0.31	1.58
66 16	2.5	69	-1.49 ± 0.23	0.22
66 18	0.7	23	-1.11 ± 0.31	1.14
66 20	1.6	137	-1.05 ± 0.21	2.80
66 24	5.6	69	-0.42 ± 0.22	1.05

Table 2 Parameters of Abell 2218

Abell cluster centre (1950.0) ⁵	RA 16h 35min 42s Dec 66° 20'
New centre position (1950.0)	RA 16h 35min 37s ± 5s Dec 66° 18'.5 ± 0'.5
Best X-ray position (1950.0)*	RA 16h 35min 42s ± 40s Dec 66° 20' ± 4'
Distance class ⁵	6
Richness class ⁵	4
Redshift ⁷	0.207
Distance (for $H_0 = 50 \text{ km s}^{-1} \text{ Mpc}^{-1}$)	1,262 Mpc
Cluster core radius, a (defined in text)	$(1.9 \pm 0.2) \times 10^{-4} \text{ rad}$ (240 kpc at A2218)
X-ray luminosity (2–10 keV)*	$(9 \pm 5) \times 10^{36} \text{ W sr}^{-1}$
Best-fitting atmosphere for $\gamma = 5/3$, $\xi = -0.15$:	
Central density	$2 \times 10^{-24} \text{ kg m}^{-3}$
Central temperature	$8 \times 10^8 \text{ K}$
Mass of gas	10^{45} kg
(Mass of gas)/(total mass of cluster)	0.08

* I. McHardy personal communication.

oversampling at the scan centre. In view of the asymmetry of the temperature decrements about the Abell centre at 66° 20', galaxy counts of this region were made from a copy of the Palomar Observatory–National Geographic sky survey red plate. New values for the position of the cluster centre and for the core radius, a , fitted to the assumed total density distribution $\rho(r) = \rho(0)(1 + r^2/a^2)^{-3/2}$ were determined from the counts—these and other cluster parameters are given in Table 2. Strip counts across A2218 and the measured values of the temperature decrement, ΔT , are plotted together in Fig. 1. It is evident that the Zel'dovich cooling is strongest at the cluster centre and decreases rapidly away from this point. A similar scan on a region of blank sky shows no such systematic deviations³.

To evaluate the reliability of the detection of an extended dip in the background, the data of Table 1 were fitted to four models; (1) no effect—a zero value at each point in the scan; (2) a constant effect, ΔT_0 , corresponding to a systematic offset; (3) a point-like 'source' at the cluster centre, parameterised by the peak cooling, ΔT_0 , and convolved with the beamshape; (4) an extended, adiabatic atmosphere with polytropic index $\gamma = 5/3$, and cluster filling factor³ $\xi = -0.15$, again convolved to the beam and parameterised by ΔT_0 .

In the last three cases, the best-fitting values for ΔT_0 were found, and χ^2 were calculated for all the models. The results were: (1) $\chi^2 = 83.4$, 5 d.f.; (2) $\chi^2 = 20.1$, 4 d.f.; (3) $\chi^2 = 13.7$, 4 d.f.; (4) $\chi^2 = 6.4$, 4 d.f. These values are such that all but the last hypothesis can be rejected at the 1% level (the probability of a greater value of χ^2 for model 4 is about 0.18). Hence our data constitute strong evidence for an extended atmosphere in Abell 2218. The ΔT values corresponding to the best atmosphere of type 4 are plotted in Fig. 1. The central temperature and density required to produce the observed X-ray flux density and temperature decrement are given in Table 2. The values of these parameters are uncertain by a factor ~ 3 due to the range of possible values of ξ .

Because the cooling effect, X-ray emission and galaxy distribution are strongly correlated in position, we conclude that the evidence for a microwave background decrement associated with a hot intracluster medium in Abell 2218 is now overwhelming.

We thank the director of the SRC Appleton Laboratory for the facilities provided at the Chilbolton Observatory.

M. BIRKINSHAW
S. F. GULL

*Mullard Radio Astronomy Observatory,
Cavendish Laboratory, Madingley Road,
Cambridge, UK*

K. J. E. NORTHOVER
16 Barry Road, London SE 22, UK

Received 7 July; accepted 25 July 1978.

1. Gull, S. F. & Northover, K. J. E. *Nature* **263**, 572 (1976).
2. Lake, G. & Partridge, R. B. *Nature* **270**, 502 (1977).
3. Birkinshaw, M., Gull, S. F. & Northover, K. J. E. *Mon. Not. R. astr. Soc.* (in the press).
4. Sunyaev, R. A. & Zel'dovich, Ya. B. *Comments Astrophys. Space Phys.* **4**, 173 (1972).
5. Abell, G. O. *Astrophys. J. Suppl.* **3**, 211 (1958).
6. Birkinshaw, M. *Mon. Not. R. astr. Soc.* **184**, 387 (1978).
7. Leir, A. A. & van den Bergh, S. *Astrophys. J. Suppl.* **34**, 381 (1977).

Limited solubility of iron in the Sun's interior

STRIPPED iron nuclei in a hydrogen plasma under central solar conditions, according to the classical Debye–Hückel model, would undergo phase separation for concentrations well below the cosmic abundance value. The higher concentration corrections, needed to characterise the iron-rich phase, lead to enhanced solubility for a simplified model where the electrons form a uniform background. Support for an iron-rich phase coalescing in the solar interior requires more accurate treatment of bound and partially bound electrons in such a mixture. The results of the Debye–Hückel model where the electrons are treated discretely and as a continuum, are reported here and support the possibility of phase separation. The physical cause of that phase separation is simply that the potential energy is lower in the separated phases than in the mixture because the local charge neutralisation is much better satisfied in the two separated phases.

Recent phase separation calculations for mixtures of stripped atoms^{1–3} suggest that highly charged nuclei would have limited solubility in a hydrogen plasma under conditions found in stellar interiors. These results suggest that highly ionised iron would precipitate out in the solar interior and coalesce in the centre. Should this occur and lead to a reduction of the iron concentration near the thermonuclear burn region, by an order of magnitude below the cosmic abundance value (2.5×10^{-5} ionic mole fraction), the existing neutrino dilemma might be resolved. It is estimated⁴ that such a concentration reduction would lower the opacity, and hence the temperature, sufficiently to significantly reduce the probability of fusion of α particles that generates the nuclei whose neutrino emission is being monitored. This paper presents various simplified models for the iron, hydrogen plasma phase diagram which explore this possibility.

A realistic calculation of the phase diagram is difficult for two reasons. The most important difficulty is that of accurately treating the bound and partially bound electrons of iron. At the solar interior temperature of about 1.5 keV and pressure of 10^5 Mbar, the two 1s electrons (with ionisation potentials of over 8 keV) are still bound whereas the eight 2s and 2p electrons (with ionisation potentials below 2 keV) are partly bound.

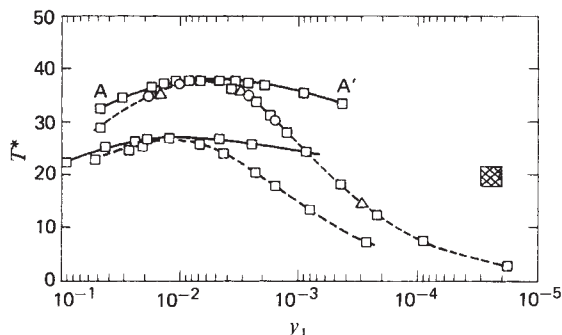


Fig. 1 The coexistence (solid lines) and instability (dashed lines) curves for an element of charge 26 (upper set of curves) and of charge 20 (lower set of curves) in a hydrogen plasma according to the Debye-Hückel theory. The ionic mole fraction of the element, y_1 , is plotted against the reduced temperature, T^* , at various pressures (Δ , 10^3 Mbar; $\square$, 5×10^4 Mbar; $\circ$, 10^4 Mbar). Points A and A' are, for example, two coexisting phases in the charge 26 system. The hatched area represents central solar conditions.

Of course, the bound-free transitions of these electrons lead to the large contribution to the opacity of a hydrogen plasma of even a very small amount of iron. This problem is crudely dealt with here by ignoring the partly bound electrons and considering the bound electrons merely to reduce the point nuclear charge to give some effective point charge. This crude approximation may not as seriously affect the calculation of the coexistence curve as might at first be expected. Errors tend to cancel as the chemical potential of iron in the iron-rich phase is equated to that in the iron-poor phase, where in both phases the electron density is nearly the same. In previous calculations³, when the polarisation of the electron gas by the ionic charge distribution was treated by perturbation theory the phase diagram was not drastically modified.

The second difficulty concerns accurate treatment of concentrated solutions of electrolytes, such as the iron-rich phase. The dilute phase can be adequately dealt with by the classical Debye-Hückel theory of a mixture of three components (iron nuclei with an effective charge, protons and electrons) as can be seen from the ratio of the ion sphere radius to the Debye length being less than unity (see Table 1). The electrons in the solar interior are very non-degenerate, as can be seen from the calculated Fermi temperature (Table 1). To avoid dealing with the concentrated solution, the instability curve was calculated in the dilute phase from the vanishing of the second derivative of the free energy with respect to iron concentration. This limit turned out, however, to be too loose to give insolubility near solar concentrations, so we had to deal with the iron-rich phase by corrections to the Debye-Hückel theory.

These corrections require a quantum mechanical treatment of the electrons to avoid divergences which arise if bound state contributions are ignored. These corrections can be included using an activity expansion⁵ but are of such complexity that they have not yet been calculated for this system. Instead, a model treating the electrons in the uniform background approximation was explored. Comparison with the discrete case discussed earlier, where in both cases the ions were treated in the Debye-Hückel approximation, was satisfactory, demonstrating the insensitivity of the coexistence curve to the electron distribution. For the uniform electron model it was then relatively easy to calculate the higher concentration⁶ corrections and show how they affect the coexistence curve. Even these corrections fail to extend the coexistence curve into the solar concentration region since the series fails to converge for the coexisting iron-rich phase.

Having exhausted so-called analytical methods, the uniform electron background model can be extended by numerical techniques using either the Monte Carlo or molecular dynamics method. These calculations are fairly cumbersome, so they will

be used initially to check how well the hypernetted chain integral equation represents the mixture of charged nuclei in a neutralising rigid uniform electron background. If that integral equation were also adequate when the charge difference between the species is extreme, as in the present case, it would reduce the numerical work required. The object of these calculations would be to determine the solubility for an iron ion of charge 24, corresponding to two bound K electrons, in a hydrogen plasma in which the polarisation of the partially bound L electrons will be treated by linear dielectric response theory.

For a mixture of N_2 protons and N_1 ions of charge Z_1 and hence $N_1 Z_1 + N_2$ classical electrons in a volume Ω , the low density value of the free energy is given by the Debye-Hückel theory⁷, which requires that the ratio of the ion sphere radius r , to the Debye length, λ , to be < 1 .

The Debye length is given by $\lambda^{-2} = 4\pi\rho e^2 \bar{Z}^2 / kT$, where $\rho = 3/4\pi r^3$ is the total particle density and defines the ion sphere radius. $\bar{Z}^2 = X_1(Z_1^2 + Z_1) + 2X_2$, where X_1 is the concentration per particle or $X_1 = N_1/N$ with $N = 2N_2 + (1 + Z_1)N_1$. We define the concentration per ion as $y_1 = N_1/(N_1 + N_2)$.

The pressure and Gibbs free energy thus derived are used to investigate phase separation at constant temperature and pressure, by the double tangent construction which is equivalent to equating chemical potentials. The instability points determined by setting the second derivative of the Gibbs free energy, with respect to mole fraction, equal to zero, always lie within the concentration range determined from the double tangency construction. The instability curve for the low concentration phase can be calculated over a wider concentration range than the phase diagram, as shown in Fig. 1, because it is not limited by the validity of the Debye-Hückel approximation for the coexisting concentrated phase. However, the applicability of the Debye-Hückel theory as given by the r/λ criterion for the free energy is severely tested when a second derivative of the free energy is taken, indicating that r/λ must be $\ll 1$ before the second derivative is adequately given by the low density theory. But as shown in Fig. 1, the instability criterion is not restrictive enough. For phase separation to occur, solar conditions must lie in the two phase region, that is below the solid curves. If solar conditions are also below the dashed curves, the instability criterion would have been sufficient.

The coexistence curve and instability curve for $Z_1 = 26$ and $Z_1 = 20$ are displayed in Fig. 1 at various pressures. The results are nearly pressure independent when plotted as a function of the reduced temperature, $T^* = kT/e^2\rho^{1/3}$, because the equation of state is almost that of an ideal gas. This is indicated by the large value of T^* which is a measure of the ratio of kinetic

Fig. 2 The coexistence curves for an element of charge 26 in a hydrogen plasma according to the Debye-Hückel theory (curve 1), according to the Debye-Hückel theory for the ions but with the electrons treated in the uniform background approximation (curve 2), according to the same model as represented by curve 2 except that the first order concentration corrections as given by Abe are included (curve 3) and according to the previous model except that the second order concentration corrections are also included (curve 4). The ionic mole fraction y_1 , is plotted against the reduced temperature, T^* , at a pressure of 5×10^4 Mbar.

Points B and B' are, for example, two coexisting phases.

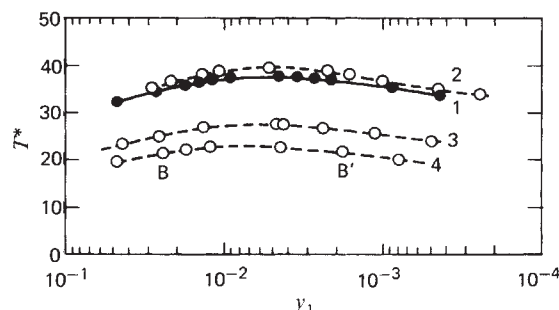


Table 1 Conditions for coexisting phases* and for the solar centre† (a) and the corresponding convergence of the Abe expansion (b)

a					
y_1	$T(10^7 \text{ K})$	$V (\text{cm}^3 \text{ mol ion}^{-1})$	P	r/λ	T/T_F
1.8×10^{-3}	1.15	0.03924	50,000	0.62	4.3
2.5×10^{-2}	1.15	0.04805	50,000	1.71	3.6
2.5×10^{-5}	1.60	0.00902	300,000	0.46	2.3
b					
y_1	P_{el}	$P_{\text{DH}}^\ddagger$	$P_{\text{Abe 1}}$	$P_{\text{Abe 2}}$	
1.8×10^{-3}	25,857	24,051	77	15	
2.5×10^{-2}	32,969	14,398	2,002	631	
2.5×10^{-5}	153,224	146,737	38	1	

* The first two entries correspond to conditions B and B' in Fig. 2, respectively.

† The third entry.

‡ The Debye-Hückel pressure includes the ideal gas term, all pressures are measured in Mbar.

phase. Nevertheless, the calculations presented here suggest phase separation under solar conditions. More realistic models are needed to confirm this suggestion.

We thank F. J. Rogers for helpful discussions.

E. L. POLLOCK
B. J. ALDER

Lawrence Livermore Laboratory,
University of California,
Livermore, California 94550

Received 25 April; accepted 16 June 1978.

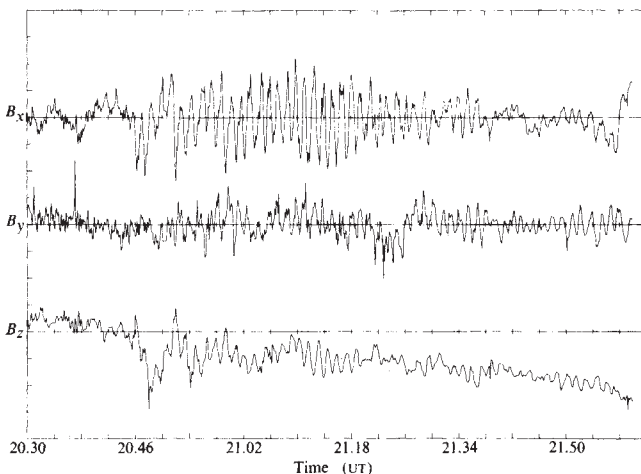
1. Stevenson, D. J. *Phys. Rev. B* **12**, 3999 (1975).
2. Pollock, E. J. & Alder, B. J. *Phys. Rev. A* **15**, 1263 (1977).
3. Hansen, J. P., Torrie, G. M. & Vieillefosse, P. *Phys. Rev. A* (in the press).
4. Bahcall, J. N. & Ulrich, R. K. *Astrophys. J.* **170**, 593 (1971).
5. Rogers, F. J. & DeWitt, H. E. *Phys. Rev. A* **8**, 1061 (1973).
6. Abe, R. *Prog. theor. Phys.* **22**, 213 (1959).
7. Brush, S. G. *Progress in High Temperature Physics and Chemistry* (ed. Rouse, C. A.), 1 (Pergamon, Oxford, 1966).
8. Clayton, D. D. *Principles of Stellar Evolution and Nucleosynthesis* (McGraw-Hill, New York, 1968).

Alfvén waves generated by an inverted plasma energy distribution

OBSERVATIONS from the geosynchronous spacecraft ATS6 of magnetic oscillations transverse to the Earth's magnetic field in the presence of a plasma ion distribution with a 'bump-in-tail' form are reported here. We argue that the oscillations are Alfvén waves, that the non-monotonicity (that is, population inversion) of the energy distribution provides a source of free energy for the oscillations, and that the particle-wave interaction is through bounce resonance¹. The data reported here comes from the UCLA fluxgate magnetometer² and the UCSD particle detector which measures ions and electrons with energies from 10 eV to 80 keV (ref. 3). The magnetometer data are displayed in a field aligned coordinate system in Fig. 1. A wave with an amplitude of about 5 nT and a period of about 100 s is present. The strongest signal is in B_x , the component in the meridian and perpendicular to the background field, B . The magnetic signal seems typical of a class of pulsations with short wavelengths across B seen in the local afternoon on ATS6⁴. These particular observations were made near 1500 LT.

Figure 2 shows the ion-velocity distribution, f , as a function of energy, W . The curves show the mean distribution for

Fig. 1 The magnetic field data from the geostationary spacecraft ATS6, are presented in a mean field aligned coordinate system in which B_x is measured along the mean field direction, obtained by forming a 10 min sliding average of the data. B_x is in the plane containing the Earth-satellite line and B_z . The B_z baseline is 130 nT. One division on the ordinate is 2 nT. Peak to peak amplitudes in B_x are of order 5 nT.



to potential energy. The solar conditions in Fig. 1 indicate that extending the coexistence curve would predict considerably smaller solubility than the cosmic abundance value of 2.5×10^{-5} in which the iron-rich phase would contain about half iron and half hydrogen. Even for the extreme case of $Z_1 = 20$, believed to represent the lowest possible ionisation state of iron under interior solar conditions, phase separation might be marginally inferred. The Debye-Hückel theory predicts, as expected, that decreasing the effective iron charge lowers the critical temperature and increases the critical concentration. The prediction is, therefore that in any stellar interior in which the difference in charge between iron and any other charged species is >20 , iron will be less soluble than its cosmic abundance. In the Sun the temperature drops at midradius sufficiently for iron ionisation to be reduced below 20 and hence beyond this midpoint, iron would be expected to be soluble.

Because the dilute Debye-Hückel theory is insufficient and the rigorous corrections require the electrons to be dealt with quantum mechanically, a uniform background electron gas approximation was explored for which concentration corrections can be more easily calculated. In that case, the free energies of the ions and electrons are additive. The electrons are treated as an ideal Fermi gas whose thermodynamic properties involve the usual Fermi-Dirac integrals⁵. For the temperature and pressure of interest here, the electron gas is nearly classical. The ions are first treated in the Debye-Hückel approximation, and the case for $Z_1 = 26$ is compared in Fig. 2 with the earlier calculation where the electrons were treated as an ionic species. The similar coexistence curve encourages the uniform background model to be extended to higher densities.

The straightforward inclusion of the first order correction^{5,6} leads to a rather large shift of the coexistence curve as shown in Fig. 2. Hence, it was necessary to consider the second order correction. The coexistence curve calculated including this correction to the free energy, is the lowest curve in Fig. 2 and shows that the shift in the critical temperature is considerably smaller than with the first correction. Extrapolating this series from these limited results leads to an estimated critical T^* , for the uniform electron gas case of about 20. This T^* is larger than that for central solar conditions, which is necessary for insolubility exceeding cosmic abundance.

Table 1 shows the convergence of the series for the pressure, at pressures corresponding to the region of thermonuclear burn and at the centre of the Sun. In the dilute iron phase, the electron and ion pressure are nearly equal, corresponding to ideal gas conditions of a hydrogen plasma, and the series convergence is excellent. In the coexisting iron-rich phase, the electron pressure exceeds the ion pressure due to the large number of electrons ionised from a high Z element. Furthermore, the convergence of the series is such that it is not possible to calculate the coexistence curve reliably for a more iron-rich

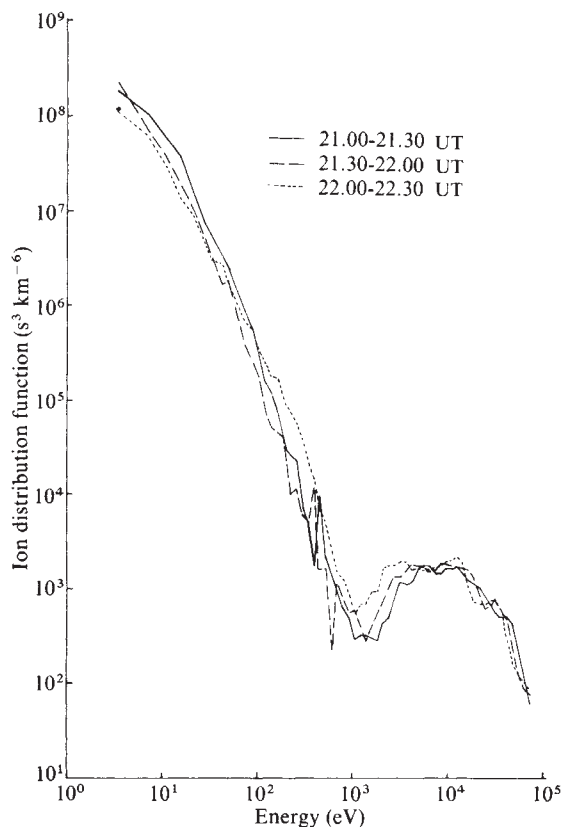


Fig. 2 Ion velocity distribution functions measured on ATS6 during the three half-hour periods during which the magnetic oscillations were observed. Although the instrument cannot distinguish ion species directly, we assume that the ions are protons.

three successive half-hour segments. The distribution remained largely unchanged through the course of the event. There is a minimum where $W \sim 1$ keV and $\partial f / \partial W > 0$ in the range $1 \text{ keV} < W < 10 \text{ keV}$. Any 'filling' of the distribution with time is probably due to spatial structure in the ion distribution, since the ions are drifting in the opposite sense to the spacecraft motion in this region⁵. In a collisionless plasma, diffusion in energy to fill the minimum may be accomplished by particles being resonantly scattered by a plasma wave. The observed frequency is of the order of the frequency of the bounce motion of an ion with a few keV energy in the Earth's mirror field and so the relevant resonant interaction is bounce resonance¹.

The low frequency and the quasi-transverse nature of the magnetic signal suggest that the wave present is an Alfvén wave with standing structure along the geomagnetic field. This idea is confirmed by more detailed investigation. The UCSD detector measures a complete energy spectrum every 24 s, and the effects of the wave can be detected in the low energy proton channels. The variation in low energy proton count is due to a drift velocity proportional to the wave electric field⁶ and using procedures similar to those used by Hughes *et al.*⁷ we found the wave electric field was in quadrature with the magnetic signal thus confirming the standing nature of the wave. The relative phase also showed that the standing wave was an even harmonic (a mode with an electric field node at the equator) and the observed frequency together with an estimate of the Alfvén speed, based on observed densities, suggests it should be the second harmonic. The symmetry of the signal about the equator is crucial in suggesting bounce resonance is the source; the first descriptions¹ of the mechanism pointed out that even harmonics would be generated. Other hydromagnetic wave sources are likely to generate the fundamental. The wavelength of the signal across the field is short, since magnetic pulsation activity on the geosynchronous SMS1 spacecraft (19° E of ATS6) and SMS2 (10° W of ATS6) did not correlate with the

ATS6 signal. This shows signals are relatively localised and rules out the Kelvin-Helmholtz instability as the energy source⁴. Spectra and coherence information from the three spacecraft for this event are shown in Fig. 4 of ref. 4.

The condition for a particle with bounce frequency, ω_b , and E-W drift speed $\tilde{\omega}_d$ to be in bounce resonance with a wave with frequency ω and E-W angular wave number m is

$$\omega - m\tilde{\omega}_d = N\omega_b$$

where N is an integer¹. Here we are interested in $N \sim 1$ and if the Doppler shift $m\tilde{\omega}_d$ is significant the E-W wavelength needs to be relatively small¹ (of the order of the equatorial Larmor radius of a resonant particle).

A particle's energy in an Alfvén wave increases due to the wave electric field at a rate $q\mathbf{E} \cdot \mathbf{v}_d$, where $\mathbf{v}_d$ is the E-W drift velocity due to curvature and ∇B drifts. We can estimate the energy diffusion coefficient due to this acceleration for a particle in resonance with a wave of bandwidth, $\Delta\omega$,

$$\langle q\mathbf{E} \cdot \mathbf{v}_d \rangle^2 \Delta\omega^{-1}$$

where the brackets denote the mean value seen over a particle bounce. We can estimate the wave electric field as $\sim 5 \times 10^{-1} \text{ mV m}^{-1}$ from the data and substitute this value into the diffusion coefficient to show that the time scale for diffusion to fill the hole in the energy distributions is $\sim 5 \times 10^5 \text{ s}$ that is a time much longer than the time scale of particle motions through this part of the magnetosphere.

The filling of the distribution seen in Fig. 2 cannot, therefore, be due to the waves. Cowley⁵ has shown how the magnetospheric convection pattern can lead to such distributions and, as indicated earlier, the protons are drifting in the opposite sense to the spacecraft motion. Much as one can estimate a diffusion time one can parametrise the energy exchange between particles and waves by estimating the contribution to the growth rate of the resonant ions. Estimates⁸ show that this is proportional to the resonant particle pressure:

$$\gamma/\omega \approx P_{\text{res}}/B^2/2\mu_0$$

which gives

$$\gamma/\omega \sim 10^{-2}, 10^{-3}$$

for observed parameters. These figures are of a similar order to the expected rates of ionospheric damping on the dayside of the magnetosphere⁹ and it seems the quasi-steady signal is maintained by energy being fed from particles to the wave to be lost in the ionosphere. The amplitude of the waves would then be sustained by the supply of ions drifting from the night side.

There are also low energy electrons in bounce resonance with the wave. Although there are more electrons than ions, their pressure is much lower as the resonance condition implies they have the same velocity as the ions.

In conclusion these observations show the first direct evidence of generation of hydromagnetic waves through bounce resonance with a non-equilibrium distribution, which theory first suggested 10 years ago^{1,10}. In this observation the distribution of ions was non-monotonic with energy. Theory shows that this is not a necessity; steep spatial gradients in the ion distribution can similarly drive waves. However, both observation¹¹ and theory⁵ indicate the regular occurrence of non-monotonic ion distributions in the afternoon-evening magnetosphere. These may be the source of many of the pc4 pulsations characteristic of this sector at geosynchronous orbit.

W. J. HUGHES

D. J. SOUTHWOOD

Blackett Laboratory, Imperial College,
London SW7, UK

B. MAUK*

Department of Physics,
University of California,
La Jolla, California 92037

R. L. MCPHERRON

Institute of Geophysics and Planetary Physics,
University of California,
Los Angeles, California 90045

J. N. BARFIELD

Space Environment Laboratory,
NOAA/ERL, Boulder, Colorado

Received 28 June; accepted 10 July 1978.

*Present address: Geophysics Program, University of Washington, Seattle, Washington 98105.

1. Southwood, D. J., Dungey, J. W. & Etherington, R. J. *Planet. Space Sci.* **17**, 349 (1969).
2. McPherron, R. L., Coleman, P. J. & Snare, R. C. *IEEE Trans. Aerosp. Electron. Syst.* **AES-11**, 1110 (1975).
3. Mauk, B. H. & McIlwain, C. E. *IEEE Trans. Aerosp. Electron. Syst.* **AES-11**, 1118 (1975).
4. Hughes, W. J., McPherron, R. L. & Barfield, J. N. *J. geophys. Res.* **83**, 1109 (1978).
5. Cowley, S. W. H. in *Physics of Solar Planetary Environments* Vol 2 (ed. Williams, D. J.) 582 (Am. Geophys. Union, Washington, D.C., 1976).
6. Kivelson, M. G. *J. atmos. terr. Phys.* **38**, 1115 (1976).
7. Hughes, W. J., McPherron, R. L., Barfield, J. N. & Mauk, B. H. *EOS Trans. Am. geophys. Un.* **56**, 723 (abstr.) (1977).
8. Southwood, D. J. *J. geophys. Res.* **81**, 4508 (1976).
9. Newton, R. S., Southwood, D. J. & Hughes, W. J. *Planet. Space Sci.* **26**, 201 (1978).
10. Southwood, D. J., Etherington, R. J. & Dungey, J. W. *Nature* **219**, 56 (1968).
11. De Forest, S. E. & McIlwain, C. E. *J. geophys. Res.* **76**, 3587 (1971).

A fast interferometric chopper for neutrons and X rays

FAST neutron choppers have been considerably developed in the past few years. The commonly used choppers are either rotating massive objects through which appropriately shaped neutron channels are cut¹ or crystals whose diffracting power may be modulated by magnetic fields² or dynamic strains³. All these devices suffer from practical limitations of speed, poor duty cycle and the consequent loss of intensity. We describe here our use of a scanning Bragg reflecting crystal interferometer as a fast chopper. Although they have been used with X rays, crystal interferometers work equally well with neutrons⁴. The interferometer is used as a phase sensitive neutron switch, in which a crystal translation of 0.96 Å is sufficient to redirect the beam. At larger translational amplitudes the interferometer has a corresponding multiplex advantage over other neutron choppers. Two of these interferometric neutron choppers could be used to construct a Fourier transform neutron spectrometer with 1,000

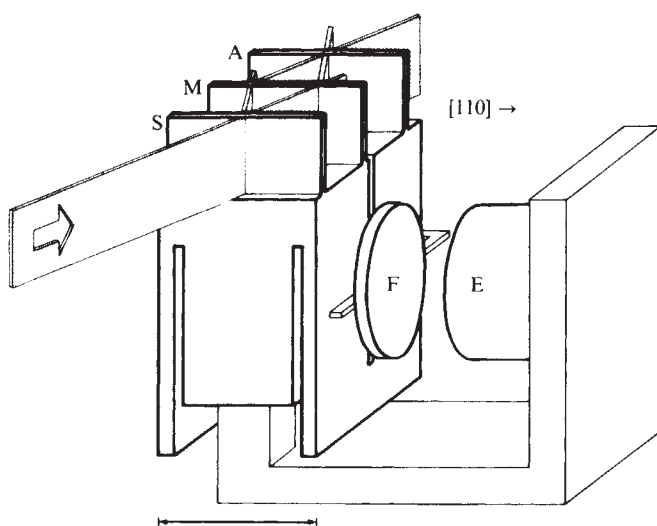


Fig. 1 The monolithic scanning X-ray interferometer which is cut from a perfect single crystal of silicon. The ferrite disk F is glued to a U-bracket which drives the elastic transverse stage and forces are generated by the electromagnet E. X-ray beam paths through the interferometer are indicated by shaded areas. 220 Bragg reflections provide beam steering in the splitter crystal S, mirrors M and analyser A. The interferometer is described in detail elsewhere⁵. Scale, 10 mm.

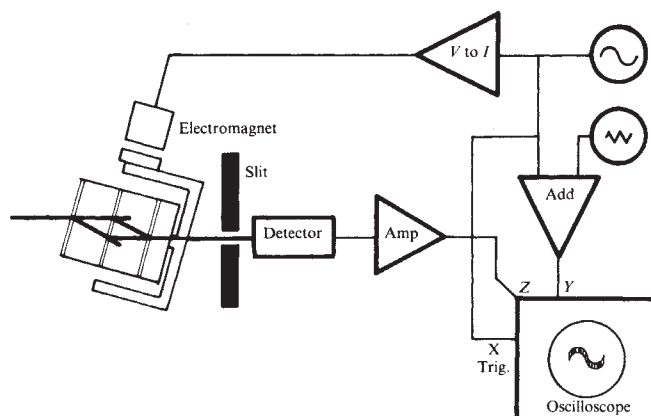


Fig. 2 The electronic drive system showing how the oscilloscope display is generated.

times the resolution of competing spectrometers. With a pulsed source a single interferometric chopper could be used to construct a time of flight spectrometer which would provide high resolution even with fast neutrons.

The interferometer construction is shown in Fig. 1. The monolithic scanning interferometer^{5,6} is driven by the force between the electromagnet E and the ferrite disk F. The three wafers S, M, A which form the X-ray or neutron interferometer, are all part of the same monolithic perfect crystal. Coherent beam steering is achieved with the 220 and $\bar{2}\bar{2}0$ Bragg reflections indicated in the ray diagram. A is connected to S and M through the monolithic elastic traverse^{5,7} which allows the translational degree of freedom of A parallel to the [110] direction. When the analyser A travels a distance of 1.92 Å in the [110] direction which is normal to the Bragg planes, one fringe is observed in each of the two output beams.

In the case of zero absorption the output beams are complementary (conservation of photons or neutrons in the analyser A). For example, when the analyser translation x is pd the forward beam T has maximum intensity, whereas for $x = (p + \frac{1}{2})d$ the intensity maximum is in the deviated R-beam. Here, p is an integer and d is the Bragg spacing; 1.92 Å for the 220 Bragg reflection from silicon. The intensity N is given by

$$N = N_0 + N_1 \sin(2\pi x/d)$$

If $x = 2 \text{ mm s}^{-1}$, which is easily achieved, then we would observe complementary output beams, amplitude modulated at 10 MHz, and with 100% duty cycle as no neutrons are absorbed in the interferometer. In practice, the fringe contrast observed with X rays may be 0.95 while for neutrons^{4,8} the best contrast observed has been approximately 0.85. Both would be 1.00 in an ideally perfect interferometer with zero absorption.

In considering the design and use of mechanical systems, one needs to know the system response to disturbances. The intensities available from both X-ray and neutron sources are so low that we could not measure the impulse response directly. We were, however, able to measure the simple harmonic frequency response directly with X rays using 30-s integration time.

Figure 2 shows the experimental arrangement and Fig. 3 demonstrates the fringes observed near a mechanical resonance in the system at 10.8 kHz. The fringe resolution is limited to about 500 ns by the time base trigger-jitter and by the frequency response of the oscilloscope z-modulation. When operating near the fundamental translational resonance which occurs at 1,420 Hz we found no loss of fringe contrast, which indicates that only the pure translational mode was excited.

The interferometer was tested with molybdenum $K\alpha_1$ radiation for which the available intensity from a conventional laboratory source is a few thousand photons per second. Each photon detected (Fig. 2) produces a bright spot on the oscilloscope screen. With 3,000 speed Polaroid film it was necessary to pulse the z-modulation for 500 ns to record an image and this

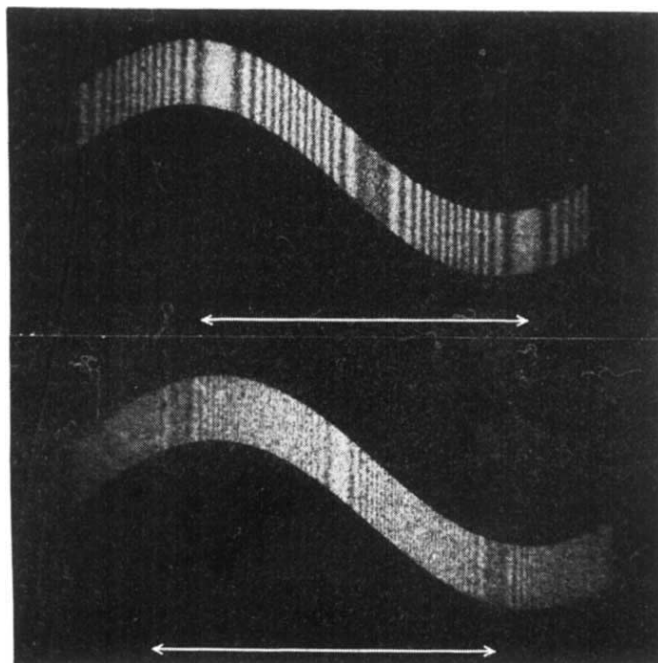


Fig. 3 X-ray fringes produced by the system outlined in Fig. 2. The finest fringes correspond to 500 ns chopping times and resolution is mainly limited by the oscilloscope performance. Each bright spot in the photograph corresponds to one detected X-ray photon. Scale, 100 μ s.

limits the present time resolution of the detector system. In Fig. 3, the interferometer was driven with a sinusoidal current of frequency ω . In a practical neutron chopper pulsed or triangular waveforms might also be used. So that the fringe system is easily visualised, the y -displacement consists of the sum of the electromagnet drive voltage and an incommensurate triangular voltage. The observed fringe spacing is, of course, not constant. The applied force is proportional to the square of the magnet current, and the observed fringes show the displacement as a function of time. The force F is given by

$$F = KI_0^2 \sin^2 \omega t = \frac{1}{2} KI_0^2 (1 + \sin 2\omega t)$$

where K is a constant and the force repetition rate is seen to be twice that of the driving current. This is clearly seen in Fig. 3 where the periodicity of the fringe pattern is twice that of the current signal. Since one fringe occurs for a displacement of 1.92 Å we were able to trace the complete frequency response of the chopper from these photographs. Thus, amplitude, stored energy and phase shift can all be measured as functions of frequency. The two fringe patterns were obtained near the 10.8 kHz mechanical resonance with only 15 mW of magnet power. The phase shift between the mechanical motion and the electromagnetic force near resonance can be clearly seen as can the ~ 500 ns spacing of the finest visible fringes. At that point the speed of the analyser wafer is 1.92 Å per 500 ns or 0.4 mm s⁻¹.

It is not unreasonable to use up 1,000 times more power in the driving electromagnet. With only 15 mW of power, the present chopper is already faster than any alternative. There are two ways in which the potential performance can be exploited. Either high power interferometric neutron chopping can be achieved on the sub-ns timescale or very precise neutron gating could be attempted on the μ s and longer timescales. The former would be attractive for extreme resolution multiplexed neutron spectroscopy using high flux neutron sources; the latter might be more attractive at pulsed sources, such as the spallation neutron source.

In future developments it will be desirable to have both positive and negative force generators. Experiments using double magnet systems and permanent magnets in place of the ferrite disk are in progress. Both systems will permit push-pull operation. The strain gradients inherent in systems which rely on

Bragg reflections from sets of standing or travelling ultrasonic waves³ have caused great problems. Because of the neutron samples' spatial and/or temporal crystal inhomogeneities, these systems are limited to frequencies well below 100 kHz with crystal thicknesses of the order of 1 mm. Provided that the interferometer wafer S, M and A undergo pure uniform translation there is no similar limit on their performance as each crystal component remains homogeneous and strain-free. The neutrons are, of course, Doppler shifted in the analyser and this is equivalent to rotating the analyser. At 2 mm s⁻¹ the rotation for thermal neutrons is ~ 0.5 arc s but the Bragg angle 2θ is not changed to first order. Since the range of Bragg reflection is ~ 2 arc s at 2 Å wavelength, the chopper could be 'returned' by rotating the analyser. At much higher crystal speeds it may become necessary to correct for the second order Doppler shift in the Bragg angle by changing the temperature of the analyser crystal. Higher order Bragg reflections can be used to provide multiplexed channelled spectra and to achieve even faster chopping rates.

At present the best on-off ratio for neutrons^{4,8} is at least 6 but with more perfect crystals this should exceed 10. Although the new chopper might have advantages in experiments to determine h/M for the neutron⁹, and in providing precisely sinusoidal neutron modulation in Fourier time-of-flight spectrometers¹⁰; other possible applications are still only in the discussion stage.

We thank Drs Hauer, Mikula, Michalec and Stirling, and the staff of the Rutherford Laboratory for helpful discussions.

M. HART

D. P. SIDONS

Wheatstone Laboratory,
University of London, King's College,
Strand, London WC2, UK

Received 17 May; accepted 21 July 1978.

1. Feshback, H. & Sheldon, E. *Phys. Today* **30**, 40-51 (1977).
2. Badurek, G. & Westphal, G. P. *Nucl. Instrum. Meth.* **128**, 315-321 (1975).
3. Mikula, P., Michalec, R. T. & Vavra, J. *Nucl. Instrum. Meth.* **143**, 23-27 (1977).
4. Rauch, H., Treimer, W. & Bonse, U. *Phys. Lett.* **47a**, 369-371 (1974).
5. Hart, M. J. *Phys. D* **1**, 1405-1408 (1968).
6. Cusatis, C. & Hart, M. *Proc. R. Soc. A* **354**, 291-302 (1977).
7. Jones, R. V. *J. scient. Instrum.* **28**, 38-41 (1951).
8. Treimer, W. *Workshop on Neutron Interferometry*, Grenoble (1978).
9. Weirauch, W. *Nucl. Instrum. Meth.* **131**, 111-117 (1975).
10. McCarthy, J. P. Jr. *Nucl. Instrum. Meth.* **121**, 609-610 (1974).

Hydrogen bond dynamics in solution

THE key to understanding the broad but featureless $\nu_s(XH)$ absorption bands of hydrogen-bonded species in solution has been the recognition by Bratos that the $\nu_\sigma(XH \cdots Y)$ hydrogen bond stretching mode has a pronounced stochastic character, and that the $\nu_s(XH)$ mode rapidly loses its phase coherence as a result of its strong anharmonic coupling to the $\nu_\sigma(XH \cdots Y)$ mode¹. This idea has been further developed by Yarwood and Robertson^{2,3}, who treat the dynamics of the ν_σ mode more precisely than does Bratos. The most detailed prediction of these theories is that the amplitude of modulation of the $\nu_s(XH)$ mode should be proportional to the r.m.s. amplitude of the ν_σ displacement coordinate r_2 , and should thus vary with temperature as $(\coth(\hbar\omega_2/2kT))^{1/2}$, where $\omega_2(=2\pi\bar{\nu}_\sigma)$ is the angular frequency of the ν_σ mode. We report here an experimental confirmation of this prediction for a dilute solution of self-associated CD₃OH in freon 11. We also investigate how the dynamics of the hydrogen bond vibration change as the solvent freezes, and we present evidence that at liquid helium temperature the $\nu_\sigma(XH \cdots Y)$ mode can be regarded as a quantum-mechanical stochastic oscillator executing a strongly over-damped zero-point motion.

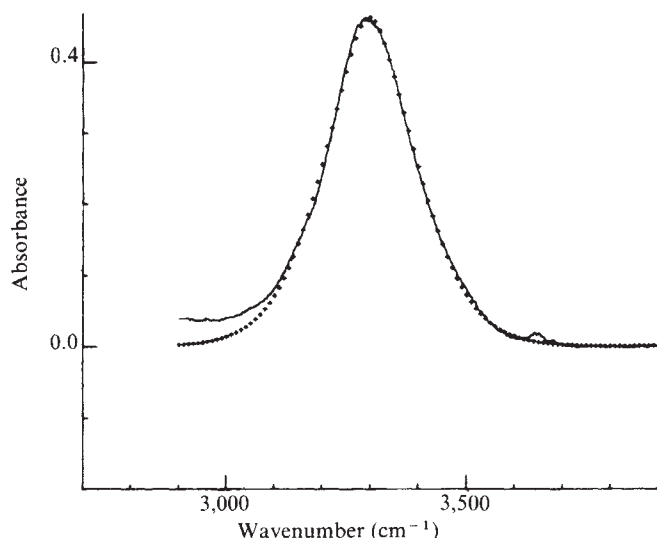


Fig. 1 Observed (continuous line) and fitted (crosses) spectrum of self-associated CD_3OH in CFCl_3 : temperature, 240 K; concentration, 0.2 M. There is a further broad absorption band on the low frequency side, not accounted for by the theory, which may be due to multimeric species. This makes it difficult to determine the low frequency baseline satisfactorily. The absorption band at $3,650\text{ cm}^{-1}$ is due to free CD_3OH ; its intensity increases with increasing temperature. The data were zero-weighted below $3,100\text{ cm}^{-1}$ and between $3,625$ and $3,675\text{ cm}^{-1}$ during curve-fitting.

Yarwood and Robertson^{2,3} assume that the ν_s displacement coordinate r_2 satisfies the Ornstein–Uhlenbeck stochastic differential equation⁴

$$\ddot{r}_2 + \gamma \dot{r}_2 + \omega_2^2 r_2 = m_2^{-1} F(t) \quad (1)$$

and, by applying Kubo's theory of a randomly modulated oscillator⁵, show that the autocorrelation function of the $\nu_s(\text{XH})$ transition dipole moment is

$$\phi(t) = \exp \left\{ -\Delta^2 \int_0^t (t-x) \psi(x) dx \right\} \quad (2)$$

where ψ is the autocorrelation function of the Ornstein–Uhlenbeck stochastic process⁴. The amplitude of modulation Δ of the ν_s mode is given by

$$\Delta = K_{112} \langle r_2^2 \rangle^{1/2} / m_1 \omega_1 \quad (3)$$

where K_{112} is an anharmonic coupling constant, and the characteristic time τ_c of the modulation is given by

$$\tau_c = \gamma / \omega_2^2 \quad (4)$$

Equations (2)–(4) give a good account of the experimental $\nu_s(\text{XH})$ absorption band profiles of phenol–acetonitrile and phenol–dioxan complexes in solution in carbon tetrachloride, correctly predict the changes which occur on deuteration, and enable the effect on the bandshape of changing the solvent to be understood⁶.

The most stringent way to test the theory, however, is to study the temperature dependence of the $\nu_s(\text{XH})$ bandshape⁷. According to equations (3) and (4), Δ should vary as $(\coth(\hbar\omega_2/2kT))^{1/2}$, while τ_c , being proportional to the microscopic viscous damping constant γ , should increase rapidly as the temperature is reduced. To investigate this, we have recorded digitally the $\nu_s(\text{XH})$ absorption bands of self-associated CD_3OH in freon 11 (CFCl_3) between 280 and 160 K, and have estimated the parameters Δ and τ_c by a nonlinear least-squares analysis of the bandshape. A typical $\nu_s(\text{XH})$ band with its theoretical reconstruction superimposed is shown in Fig. 1, and a summary of the parameter estimates is given in Table 1. The method of analysis was similar to that described previously⁶,

except that $\bar{\nu}_s (= \omega_2/2\pi c)$ was treated as a fixed parameter with a value of 100 cm^{-1} . This is a compromise between the value of 80 cm^{-1} (or at most 100 cm^{-1}) suggested by Lake and Thompson⁸, and that of 120 cm^{-1} found by Zelsmann (personal communication).

The regression $\Delta = a + b (\coth(\hbar\omega_2/2kT))^{1/2}$ is close to linear, with a correlation coefficient of 0.99. The intercept a does not differ significantly from zero at the 95% confidence level. This strongly supports equation (3), and shows that the contribution to the broadening from mechanisms other than vibrational phase relaxation is small. As expected, the viscous damping constant γ increases with decreasing temperature; however, the temperature dependence of γ cannot be well represented by a law of the form $\exp(A/kT)$, particularly at the lower temperatures. (Although this may be partly due to the uncertainty in the value of ω_2 , previous work⁶ has shown that γ is not related to the macroscopic viscosity of the solvent.) The Kubo parameter $\tau_c\Delta$ increases somewhat as the temperature is reduced, but the system is far from the slow modulation limit ($\tau_c\Delta \gg 1$) at all temperatures above 160 K, and the $\nu_s(\text{XH})$ absorption band profiles differ markedly from gaussian shape.

Below 160 K the solution freezes to form a vitreous solid. On freezing, the bandshape changes dramatically: the halfwidth at 170 K is 175 cm^{-1} , and at 150 K it is 210 cm^{-1} (see Fig. 2). However, analysis of the bandshape shows that the amplitude of modulation continues to decrease, in a manner consistent with equation (3), as the temperature is reduced to 150 K: there is no discontinuous change in the amplitude of modulation at the transition point, even though the $\bar{\nu}_s(\text{XH})$ frequency abruptly shifts upwards by several wavenumbers on freezing. (The general trend is for the $\nu_s(\text{XH})$ frequency to shift steadily downwards with decreasing temperature.) On the other hand, τ_c increases from 0.046 ps to 0.40 ps, $\tau_c\Delta$ from 0.83 to 7.0, and $\gamma/2\omega_2$ from 0.44 to 3.9.

Although these estimates are imprecise, particularly at the lower temperature where the errors are of the order of 10%, they do show qualitatively how the dynamics of the hydrogen bond change as the solvent freezes. In the solid the $\nu_s(\text{XH} \cdots \text{Y})$ vibration is heavily overdamped, and relaxes slowly. The large value of the Kubo parameter $\tau_c\Delta$ indicates that the random frequency modulation is quasi-static (that is, that $\phi(t)$ relaxes in a time τ_r much shorter than the characteristic time τ_c of the modulation), so that the $\nu_s(\text{XH})$ absorption band is nearly gaussian in shape. The spectacular increase in halfwidth on freezing can be entirely attributed to a decrease in the kurtosis of the spectrum on entering the slow modulation regime.

On further cooling to 6 K the $\nu_s(\text{XH})$ band remains gaussian in profile and the halfwidth decreases to 190 cm^{-1} ; the amplitude of modulation at 6 K is estimated by analysis of the spectrum to be 81 cm^{-1} . Although this is significantly greater

Table 1 Bandshape parameters for self-associated CD_3OH in CFCl_3 (0.2 M solution)

T/K	$(2\pi c)^{-1} \Delta\text{ cm}^{-1}$	$\tau_c\Delta$	$\gamma/2\omega_2$	$[\coth(\hbar\omega_2/2kT)]^{1/2}$
280	120.9	0.60	0.25	1.99
270	118.3	0.63	0.26	1.96
260	115.9	0.64	0.28	1.92
250	114.0	0.68	0.30	1.89
240	111.7	0.70	0.31	1.85
230	110.1	0.72	0.33	1.82
220	105.7	0.78	0.37	1.78
210	101.8	0.77	0.38	1.74
200	99.9	0.79	0.40	1.70
190	98.6	0.80	0.41	1.66
180	96.8	0.82	0.43	1.62
170	95.4	0.83	0.44	1.58
160	94.2	0.84	0.45	1.54
150	92.9	7.0	3.9	1.50
6	80.6	8.0	5.0	1.00

The figures in the last two columns assume that ω_2 is known, and are subject to possible systematic error.

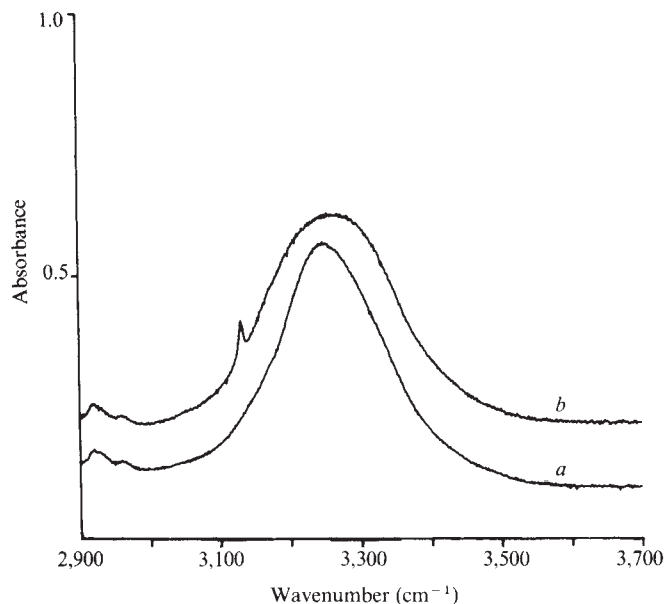


Fig. 2 A comparison of the $\nu_s(XH)$ absorption spectra at 170 K (liquid) (a) and 150 K (solid) (b). The halfwidths are 175 cm^{-1} and 210 cm^{-1} , respectively, but the amplitudes of modulation are 95 and 93 cm^{-1} . The spectrum of the solid solution is very nearly gaussian in profile. The feature at $3,130\text{ cm}^{-1}$ is an uncompensated solvent band, which is narrower and more prominent in the solid. The curves have been offset for display.

than the 95% confidence estimate $58 \pm 6\text{ cm}^{-1}$ obtained by extrapolating to zero temperature the relationship

$$\Delta = a + b[\coth(\hbar\omega_2/2kT)]^{1/2} \quad (5)$$

with parameters $a = (-4 \pm 10)\text{ cm}^{-1}$ and $b = (62 \pm 6)\text{ cm}^{-1}$ determined from measurements on the liquid between 160 K and 280 K, at least part of the discrepancy may be due to error in ω_2 . At 6 K the random frequency modulation of the $\nu_s(XH)$ band is largely produced by the zero-point motion of the $\nu_o(XH \cdots Y)$ mode.

Very similar results, consistent with equations (3) and (5), have been obtained by analysing the spectrum of a 0.2 M solution of CH_3OD in freon 11 over the same temperature range. The ratio $b(H)/b(D)$ (see equation (5)) is found to be approximately $\sqrt{2}$, in agreement with theory. However, the transmission window (or 'Evans hole') at about $2,460\text{ cm}^{-1}$ in the spectrum of self-associated CH_3OD complicates the analysis and leads to various uncertainties.

We believe that we have observed an unusual phenomenon: the zero-point motion of a heavily overdamped stochastic oscillator. Because of the strong damping and consequent slow relaxation of the r_2 coordinate in the frozen medium, the $\nu_s(XH)$ transition dipole moment relaxes in a time which is small compared with τ_c . The $\nu_s(XH)$ band thus reflects only the probability distribution of the r_2 displacement coordinate, which is gaussian. The halfwidth of the band at liquid helium temperature is consistent with the variance of this displacement coordinate being $\hbar/2m_2\omega_2$. It must be admitted, of course, that the theory in its present form^{2,3} is semi-classical, and not strictly valid at liquid helium temperatures. The theory could be made rigorous by simply replacing the classical Ornstein-Uhlenbeck autocorrelation function $\psi(t)$ in equation (2) by an autocorrelation function derived from a quantum-mechanical model of a brownian oscillator⁹. Because in the disordered solid phase the conditions for slow modulation are well satisfied, so that the bandshape depends only on the probability distribution and not on the dynamics of the modulation process, there seems little to be gained by such a refinement.

The broad $\nu_s(XH)$ bands of solutions of hydrogen-bonded species at low temperatures have sometimes been interpreted in terms of frozen-in disorder. Our work suggests that in dilute

solid solution this disorder should not be understood as a distribution of $XH \cdots Y$ equilibrium distances, but rather in terms of the slow relaxation of the r_2 coordinate back to its equilibrium position and the absence of any oscillatory character to its motion (see refs 1, 7). If there were a distribution of $XH \cdots Y$ bondlengths in dilute solid solution, it is improbable that this distribution would be gaussian or that it would produce a broadening of the $\nu_s(XH)$ band at liquid helium temperature consistent with equation (3). Perhaps the most convincing evidence in favour of our interpretation, however, is the absence of any discontinuity in the amplitude of modulation at the freezing point, despite the pronounced change in the shape of the absorption band.

We thank the CEA, CIES, and CSIR for financial support, and Dr H. R. Zelsmann for assistance. G.N.R. thanks Dr Y. Maréchal for his hospitality, and discussions.

J. BOURNAY

DRF-RM, Centre d'Etudes Nucléaires de Grenoble,
38041 Grenoble, France

G. N. ROBERTSON

Department of Physics,
University of Cape Town,
Rondebosch 7700, South Africa

Received 31 May; accepted 21 July 1978.

1. Bratos, S. *J. chem. Phys.* **63**, 3499-3509 (1975).
2. Yarwood, J. & Robertson, G. N. *Nature* **257**, 41-3 (1975).
3. Robertson, G. N. & Yarwood, J. *Chem. Phys.* (in the press).
4. Ming Chen Wang, & Uhlerbeck, G. E. *Rev. mod. Phys.* **17**, 323-42 (1945).
5. Kubo, R. in *Fluctuation, Relaxation and Resonance in Magnetic Systems*, (ed. ter Haar, D.) 23-68 (Oliver and Boyd, Edinburgh, 1962).
6. Yarwood, J., Ackroyd, R. & Robertson, G. N. *Chem. Phys.* (in the press).
7. Sandorfy, C. in *The Hydrogen Bond: Recent Developments* (eds Schuster, P., Zundel, G. & Sandorfy, C.) 615-653 (North-Holland, Amsterdam, 1976).
8. Lake, R. F. & Thompson, H. W. *Proc. R. Soc. A* **291**, 469-77 (1966).
9. Ullersma, P. *Physica* **32**, 27-55 (1966).

Origin of olivine subgrain boundaries in mantle peridotites

INVESTIGATION of the fabric, texture and dislocation microstructure of olivine has given much insight into the deformation history of mantle peridotites¹. Optically visible (100) subgrain boundaries (tilt or kink-band boundaries) are particularly significant in this respect. Their existence is the principal argument favouring intracrystalline glide as the mechanism of deformation² and the spacing between them (d_{100}) is considered to be a measure of the stress conditions during deformation³⁻⁶. Previous workers have argued that the sub-boundaries in olivine originated at the onset of deformation of peridotites^{1,2}, pre-dating both the recrystallised strain-free secondary grains⁷ and grain boundary development in general¹. However, re-examination of the distribution of sub-boundaries in olivine reveals features which are quite inconsistent with this interpretation. It is suggested in the present work that (100) sub-boundaries are a late feature in these peridotites and most probably have no connection with hypothetical asthenosphere flow.

The slip system activated during the development of (100) subgrains in olivine is $\{0kl\} [100]$ (ref. 8); the slip plane is usually subparallel to (010) and the slip direction is always $[100]$ (Fig. 1). Detailed thin section examination of mantle peridotites shows that the point of inside termination of a (100) sub-boundary (the point nearest to the common rotation axis of the two related subgrains; I intersection in Fig. 1) commonly coincides with a local convex feature on the grain boundary (Figs 2 and 3). The latter may be the corner of a polygonal or cusp-like re-entrant of the deformed grain or the point of maximum curvature of a smooth curvilinear boundary.

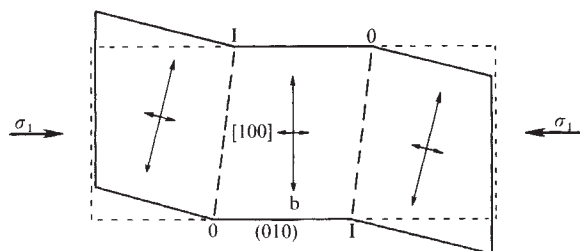


Fig. 1 Development of (100) subgrain boundaries in olivine by intracrystalline glide. (010), Glide plane; [100], glide direction; broken lines, original rectangular grain outline; σ_1 principal stress axis; I, subgrain and grain boundary intersection nearest to common rotation axis of two related subgrains.

The intersection of subgrain and grain boundaries at triple points, with reported 120° dihedral angles has been noted previously², and used as evidence for subgrain boundary development after (100) sub-boundary formation. Textural evidence contradicting this position is as follows. (1) Most olivine grains in peridotites have curvilinear to sinuous grain boundaries and re-entrants are just as cusped on grain boundaries parallel to sub-boundaries as on grain boundaries normal to sub-boundaries. 'Dihedral angles' are not in any way exclusively or even predominantly associated with the intersection of subgrain and grain boundaries. (2) The actual dihedral angles vary markedly in magnitude and show no indication of having adjusted, through recrystallisation, to equilibrium values. (3) Cusp-like re-entrants at triple points are predominantly associated with I intersections (Fig. 1) whereas re-equilibration of grain boundary energy through recrystallisation must result in a nearly equitable distribution between triple points associated with both I and O intersections (Fig. 1). (4) (100) Sub-boundaries frequently have uneven spatial distribution within individual grains, which, in itself, strongly suggests that they are localised by the convex features on the grain boundaries. (5) Sub-boundaries in primary (large) olivine grains also intersect convex features on secondary grain boundaries, and, although they are discontinuous in largely or wholly enclosed secondary grains, they do not have continuity across them. Hence, secondary olivine grains also seem to predate sub-boundary development.

Further evidence for the late development of sub-boundaries is suggested by their abundance in the elongated olivine grains of foliated porphyroclastic and equigranular peridotites⁷. The stress system required for (100) subgrain formation (Fig. 1) and the resulting deformation, shortening parallel to [100], are entirely inconsistent with the proposed deformation mode of these peridotites, translation gliding on $\{0kl\}$ [100] (refs 1, 2, 9). An attempted explanation of this inconsistency in the literature² is quite unconvincing, since the net elongation of the olivine grain under consideration is due to grain rotation, not subgrain formation. Moreover, olivine grains in these peridotites would continue to elongate with continued deformation, so that early formed sub-boundaries should become smeared and irregular. In fact, as is well known, (100) sub-boundaries are usually sharply defined and linear.

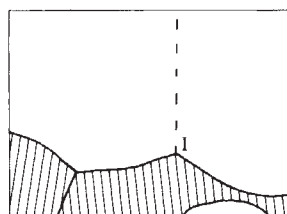
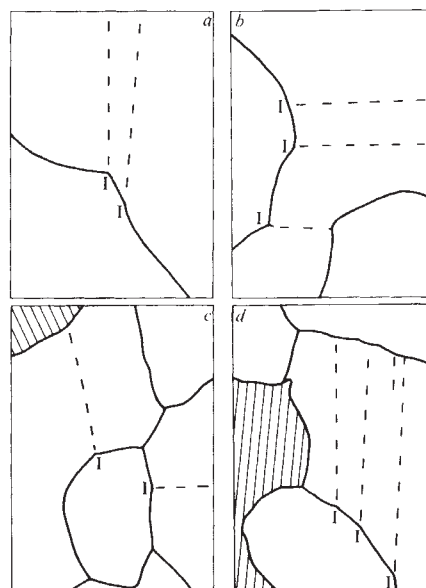


Fig. 2 (100) Sub-boundary in peridotite (spinel lherzolite) nodule from Camperdown, Victoria, Australia. Unshaded grains, olivine; shaded grains, orthopyroxene; I, I intersection (Fig. 1).

I suggest that the sub-boundaries in olivine grains of mantle peridotites developed during a later phase of deformation than that which produced the overall rock fabric. As presently envisaged, this later deformation was most probably associated with structural disruption leading to inclusion of the peridotite xenoliths in the basalt liquids. The sub-boundaries are nucleated at the convex features of appropriately orientated grain boundaries, their development being activated by the local stress concentration expected at these points. The role of stress concentration at the tip of Griffiths cracks in the brittle fracture of solids is well understood¹⁰. A direct analogy with the development of (100) sub-boundaries is the nucleation of annealing twins in face centred cubic metals¹¹. The distribution of sub-boundaries in a peridotite does not reflect the overall stress on the rock, but the local stress on individual grains. Hence, the preferred orientation of sub-boundaries in foliated peridotites is inherited from the preferred orientation of the olivine grains themselves. In undeformed peridotites the sub-boundaries are randomly orientated. The lack (or minimal development) of (100) sub-boundaries in secondary olivine grains is attributed to their smaller size and more regular and equidimensional shape, which allows them to resist deformation by subgrain formation much more readily than the larger, more irregular and highly crenulated primary grains. Similarly, partly or wholly included secondary grains appear passive to the deformation not because they postdate it but because they are inappropriately orientated to allow translation gliding on the same system as the host.

Fig. 3 (100) Sub-boundaries in Camperdown spinel lherzolite (see Fig. 2).



Recent d_{100} calculations of the stresses under which mantle peridotites have been deformed give values about one order of magnitude greater than that derived from plate tectonic models on the rheology of the mantle^{5,6}. This discrepancy has led to the suggestion that the deformation causing subgrain development occurred during diapiric upwelling of the peridotite rather than during horizontal asthenospheric flow⁵. This is essentially consistent with the present report, although the possible deformation mechanism is not necessarily limited to diapiric upwelling. Grain boundary stresses developed immediately before magma incorporation might be sufficient for subgrain development. Should this be the case, the presence of (100) sub-boundaries would not be a useful criterion to distinguish between exotic and cognate xenoliths¹². Quite clearly, the crystallisation and deformation histories of peridotites preserved as

nodules in basalts are far more complex than had been anticipated in the not-too-distant past, and the extent of preservation of asthenospheric flow characteristics in them, if indeed such ever existed, remains to be demonstrated.

I thank N. D. MacRae for rock specimens. This study was supported by the NRC of Canada.

MICHAEL E. FLEET

Department of Geology,
University of Western Ontario,
London, Ontario, Canada N6A 5B7

Received 26 May; accepted 21 July 1978.

1. Nicolas, A. & Poirier, J. P. *Crystalline Plasticity and Solid State Flow in Metamorphic Rocks* (Wiley, New York, 1976).
2. Nicolas, A., Bouchez, J. L., Boudier, F. & Mercier, J. C. *Tectonophysics* **12**, 56–86 (1971).
3. Raleigh, C. B. & Kirby, S. H. *Mineral. Soc. Am. Spec. Pap.* **3**, 113–121 (1970).
4. Goetze, C. *Geology* **3**, 172–173 (1975).
5. Gueguen, Y. *Tectonophysics* **39**, 231–254 (1977).
6. Nicolas, A. & Ricoult, D. *2nd Int. Kimberlite Conf. Abstr.* (1977).
7. Mercier, J. C. & Nicolas, A. *J. Petrol.* **16**, 454–487 (1975).
8. Raleigh, C. B. *J. geophys. Res.* **73**, 5391–5406 (1968).
9. Nicolas, A., Boudier, F. & Boullier, A. M. *Am. J. Sci.* **273**, 853–876 (1973).
10. Honeycombe, R. W. K. *The Plastic Deformation of Metals* (Edward Arnold, London, 1968).
11. Dash, S. & Brown, N. *Acta met.* **11**, 1067–1075 (1963).
12. O'Hara, M. J. in *Ultramafic and Related Rocks* (ed. Wyllie, P. J.) 346–349 (Wiley, New York, 1967).

Some new measurements on Stonehenge

THE possible astronomical significance of Stonehenge has made it desirable to make new measurements of some of its features, and I report here the results of a survey made in April 1978. New values are given for the axis of symmetry of the Avenue, the figure of the four Stations, the original position of the peak of the Heel Stone and the axis of symmetry of the sarsen trilithons (Fig. 1).

Linear measurements were taken to the nearest 5 mm with a Fibron glass-fibre tape at constant tension, calibrated against a steel tape kept as a standard. Angular measurements were made with a theodolite reading directly to 20" and by estimation to 5". The coordinate system used (Tables 1–3) has an origin 141.42 m south-west of the bronze peg β close to the centre of Stonehenge¹, with axes directed to true north and east, so that the coordinates of this peg are 100.00 E, 100.00 N. This system of local coordinates is to be preferred to National Grid coordinates, because it allows direct conversion to true azimuths. Heights are given in metres above Ordnance Datum, the level of the bench mark on stone 16 being taken as 103.11 above OD. The full National Grid coordinates of the peg β are SU 12247.43, 42195.20 (ref. 1). The referring object at the peg β was the west side of a prominent square brick chimney at Larkhill (National Grid reference 12750, 44530), the azimuth of which had been determined by Thom in 1973 as 12°10'24" (personal communication).

The part of the Avenue surveyed was the straight section which runs downhill northeastwards from the entrance of the Stonehenge earthwork. Its length to the point where it begins to change direction eastwards is 509 m. Eighteen transverse profiles were measured with a staff and level to find the lowest point of the surface over the ditch on either side, and a peg was inserted halfway between each of these pairs of points. The first of these profiles was about 2 m from the beginning of the Avenue and the last about 16 m from the end of the straight section. The remainder were spaced about 30 m apart, except where the modern road (A 344) crosses the Avenue. The coordinates of the index-marks on the pegs were then determined by a theodolite traverse. The closing error over a distance of 982 m was 0.087 m, giving a consistency of measurement of better than 1 part in 10,000. The adjusted coordinates of the 18 pegs A–T are shown in Table 1. The traverse began and ended at peg A.

Regardless of the method adopted for estimating the azimuth of the straight line which is the 'best fit' to these 18 points, the

result is extremely close to 49°54'40", and this is accordingly taken to be the azimuth of the axis of symmetry of the straight section of the Avenue. This value has been confirmed by direct observations of the azimuths of the lines joining the ends of the two Avenue ditches. At the Stonehenge end these were excavated by Hawley in 1923 (ref. 2), and were located by probing for the lowest point of the ditch bottoms in 1978. At the far end, where the Avenue begins to change direction, both ditches were exposed by excavation in 1978. The azimuth of the north-west ditch, end to end, is 49°52'19", and that of the south-east ditch 49°56'5", both values being the mean of six observations. The mean of these two values, 49°54'12", agrees well with the calculated axis.

This new value differs significantly from the mean axis of 49°35'51" found by Lockyer³, and from the axis of 49°34'18" which he adopted for reasons which can now be seen to be irrelevant. Lockyer used only six pegs on the apparent centre line of the Avenue, and did not record their positions in detail. Two of them were 140 ft apart "near the commencement of the avenue", and four others "at distances averaging 100 feet apart nearer the further recognisable extremity". Elsewhere, he states that the Avenue was discernible "for more than 1,300 ft from the centre of the temple". This implies that his furthest peg lay not more than 425 m from the centre of Stonehenge. The straight section of the Avenue continues, however, for a further 139 m beyond this point, although it has been somewhat reduced by cultivation.

It is no longer necessary to attempt to derive the date of construction of this section of the Avenue from its orientation, as Lockyer did, because there are now two closely agreeing radiocarbon dates from deer antler picks found on the bottoms of the two ditches close to Stonehenge. These are BM-1164, 1728 ± 68 b.c. and HAR-2013, 1770 ± 70 b.c. The weighted mean of these dates, when corrected⁴, is 2165 ± 80 b.c.. At this date the first gleam of the summer solstice sunrise (with 2' of the upper limb above the horizon) would have had an azimuth of 49°30', which differs from the axis now determined by about 25'.

Fig. 1 Diagrammatic plan of Stonehenge, showing the beginning of the Avenue, the Heel Stone, the four Stations (91–94), the sarsen circle and horseshoe, and the survey point β . Standing stones are shown in black, fallen and missing stones in outline.

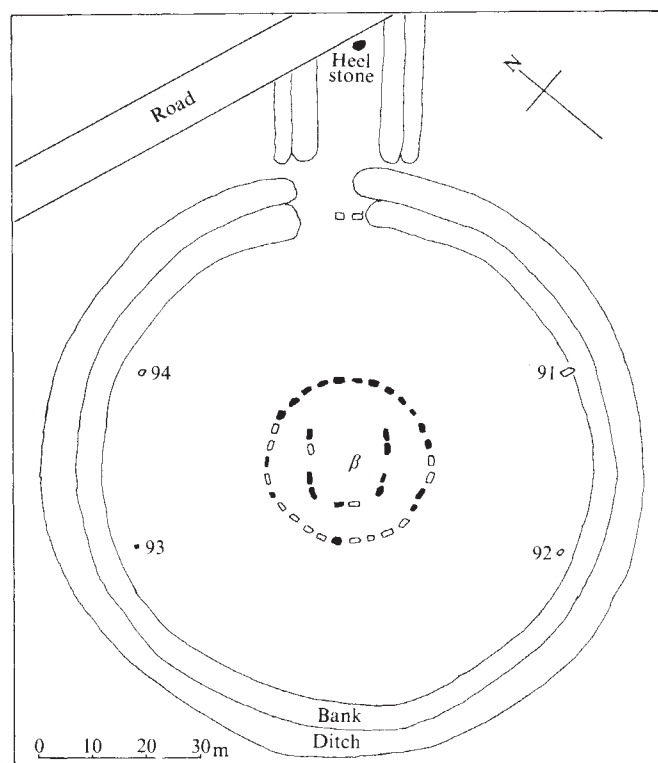


Table 1 Coordinates of Stonehenge centres and midpoints of the Avenue

Survey point	Eastings	Northings	Height
Centre of Aubrey Holes	99.58	99.37	
Centre of sarsen circle	99.15	99.61	
β	100.00	100.00	102.33
A	143.37	137.55	101.25
B	155.40	147.62	100.70
C	178.56	166.52	99.95
D	197.87	183.59	99.12
E	220.32	203.07	97.79
F	243.33	222.05	96.53
G	266.14	241.38	95.24
H	288.97	260.56	94.07
J	311.81	279.78	93.26
K	334.76	298.75	92.61
L	357.53	317.98	92.05
M	380.27	337.24	91.53
N	403.36	356.10	91.41
P	425.63	375.87	91.50
Q	448.61	394.88	91.31
R	471.65	413.80	90.96
S	494.41	433.04	90.53
T	519.00	453.54	89.60

If, however, there were trees growing on the skyline to a height of 10 m, which is by no means impossible, the azimuth of the apparent sunrise thus defined would have almost precisely matched the azimuth of the axis now determined.

If it is assumed that the lines of the Avenue ditches were laid out by lateral offsets from a central axis, and that this axis passed through the point defined by the means of the coordinates of the points A–T, then the axis prolonged southwards would have passed 1.46 m north of the peg β , 1.14 m north of the centre of the sarsen circle and 1.74 m north of the centre of the Aubrey Holes¹. The centre of the unfinished double circle of bluestones of period II (ref. 5) is not known precisely, because only a small proportion of its stoneholes have so far been excavated; however, the available evidence suggests that it was close to the centre of the subsequent sarsen circle. These divergencies imply, therefore, that the double bluestone circle and the Avenue, both of period II, did not share a common axis of symmetry.

The four Stations lie approximately on the circumference of the circle of Aubrey Holes, and are often stated to have formed a rectangle with sides aligned on extreme azimuths of the rising of the Sun and Moon⁶. Until the present survey, however, the position of Station 94 was located only by inference or conjecture. The values given here (Tables 2 and 3) are derived from direct observation and measurement on the ground.

Station 91 is represented by a fallen stone, now leaning outwards against the inner slope of the bank, with an overall length of 2.75 m. The original position of its centre is best estimated from the outline of the steep-sided stonehole in which it stood. This was found by probing in April 1978.

At Station 92, the centre of the deepest part of the stonehole can be determined from an unpublished plan made after its excavation in 1921 (ref. 7), and was confirmed by probing in July 1973 and again in April 1978.

Station 93 is occupied by the stump of an upright stone, but the dressing of parts of its sides below present ground level (and well below prehistoric ground level), like that of the sarsen stones of period IIIa, suggests that it may be a later replacement of an earlier marker. The other surviving Station stone, no. 91, shows no signs of dressing on its visible surface. The centre of stone 93 was determined by direct measurement.

The stonehole at Station 94 was located by excavation for the first time in April 1978, when the gravel access-path beneath which it lies was being replaced by turf. The filling was removed only far enough to define the upper edge of the hole, at the surface of the chalk, but probing showed that the sides descended almost vertically. Its eastern margin, where it may overlap

the adjacent Aubrey Hole 46, was left undisturbed. This Aubrey Hole was probably excavated, in part at least, by William Cunnington on behalf of Sir Richard Colt Hoare early in the nineteenth century. The latter records the finding of a cremated human burial⁸, many of which have been found in the other excavated Aubrey Holes⁹.

The coordinates of the apparent centres of the four Stations, the included angles at each, and the lengths and azimuths of the sides of the quadrilateral thus defined are given in Tables 2 and 3. These values suggest that the figure of the four Stations could have been intended to be a rectangle, with adjacent sides and diagonals roughly in the ratio of 5:12:13, the Pythagorean triangle suggested by Dibble¹⁰. From the evidence available, however, it is impossible to determine the precise shape, size or orientation of the original figure, because it cannot be assumed that the peaks of the stones (the points of reference for astronomical observations) were identical in plan with the centroids of the holes in which they stood. For this reason the original azimuths may have differed from those given here by at least $\pm 1^\circ$ for the shorter sides of the figure, and by at least $\pm 30'$ for the longer. These uncertainties must be taken into account in any calculation which is made with the aim of showing that the four Stations had an astronomical significance.

The Heel Stone now leans very markedly from the vertical, although this has sometimes been ignored in discussing its astronomical significance¹¹. By a combination of direct measurement and photogrammetry, its inclination has been determined as $24^\circ 26'$ from the vertical, in a plane bearing $233^\circ 22'$ from true north. The coordinates of its peak, as seen today from the centre of Stonehenge (peg β) are 160.21° E, 148.75° N, 105.07 m above OD. If it were set vertical, its present apparent peak would lie a little beyond and below its new apparent peak, the coordinates of which would be 161.94° E, 150.04° N, 105.60 m above OD.

The Heel Stone is part of the structure of period I, for which the best estimate of the centre is that of the circle of Aubrey Holes (Table 1). The ground levels within the stones of Stonehenge have been altered during the last 25 yr by re-turfing, by the laying of gravel in 1963 and by its replacement with turf in 1978; but extrapolation from adjacent contours on undisturbed ground suggests that in the absence of disturbance the ground level at the Aubrey centre would today be 102.38 m above OD. If an allowance of 0.56 m is made for the weathering of the chalk since the building of Stonehenge I (ref. 12), and of 1.55 m for the eye-height of a prehistoric observer¹³, the level of a prehistoric observer's eye would have been at 104.49 m above OD. From this position the azimuth of the peak of the upright Heel Stone would have been $50^\circ 54'$, and its altitude would have been $47^\circ 29'$, so that 0.3 m of its top would have projected above the bare horizon. Once more, however, if there were trees on the skyline standing more than 10 m high, the peak of the stone would not have projected above the apparent horizon for an

Table 2 Coordinates and included angles at the four Stations

Station	Eastings	Northings	Included angle
91	139.00	80.00	$89^\circ 24'.6$
92	113.03	57.80	$89^\circ 32'.5$
93	61.58	118.97	$91^\circ 2'.4$
94	86.22	140.47	$90^\circ 0'.5$
Intersection of diagonals	99.37	99.99	$45^\circ 11'.4$

Table 3 Distances and azimuths between the stations

Stations	Distance	Azimuth
91 to 92	34.17	$229^\circ 28'.5$
92 to 93	79.93	$319^\circ 56'.0$
93 to 94	32.70	$48^\circ 53'.6$
94 to 91	80.26	$138^\circ 53'.1$

observer of the summer solstice sunrise. A question posed by Hoyle thus remains unanswered¹⁴.

Thom has shown¹ that the inner faces of the outer four trilithons at ground level are a good fit to an ellipse, the major axis of which has an azimuth of $49^{\circ}57' \pm 3'$, corresponding to astronomical sunrise at 1600 ± 450 b.c. He recognised that the second, or south, trilithon (stones 53, 54) was leaning inwards, perhaps by $3''$ at ground level, but even so, his estimate of the axis of symmetry of the trilithons was based on the present ground plan of their inner faces.

The inclination of the second trilithon was estimated in the present survey to be $1^{\circ}5'$ from the vertical, which corresponds to an inward displacement at ground level of 0.10 m. If this is corrected, the major axis of the best-fit ellipse would be rotated anticlockwise through an angle of $25'-30'$, giving a revised azimuth of about $49^{\circ}30'$. This is close to the calculated azimuth of the first-gleam summer solstice sunrise (with $2'$ of the upper limb showing) in 2045 b.c., which is $49^{\circ}31'17''$. This date is the weighted mean, when corrected, of two radiocarbon dates for the Stonehenge II/IIIa transition, BM-46, 1720 ± 150 b.c. and I-2384, 1620 ± 110 b.c. The resulting inference is that the builders of the Avenue and of the sarsen structure defined sunrise by the first gleam of the upper limb of the Sun over the apparent horizon.

I thank the National Trust and their tenant, Mr J. Wort, for permission to survey and excavate on the Avenue, the Ancient Monuments Directorate, Department of the Environment, for permission and facilities for the excavation at Station 94, the Department of Civil Engineering, University College, Cardiff for the loan of a theodolite, and my wife for much patient help.

R. J. C. ATKINSON

Department of Archaeology,
University College,
Cardiff, UK

Received 25 May; accepted 26 July 1978.

1. Thom, A., Thom, A. S. & Thom, A. S. *J. hist. Astr.* **5**, 71-90 (1974).
2. Hawley, W. *Antiq. J.* **4**, 30-39 (1924); **5**, 21-36 (1925).
3. Lockyer, N. & Penrose, F. C. *Proc. R. Soc.* **69**, 137-147 (1902).
4. Clark, R. M. *Antiquity* **49**, 251-266 (1975).
5. Atkinson, R. J. C. *Stonehenge*, 58-61, 72-77, 204-206 (Penguin, Harmondsworth, 1960).
6. Atkinson, R. J. C. *J. hist. Astr.* **7**, 142-144 (1976).
7. *Ancient Monuments Directorate*, Department of the Environment, plan 123/59A.
8. Colt Hoare, R. *The Ancient History of South Wiltshire*, 145 (Miller, London, 1812).
9. Atkinson, R. J. C. *Stonehenge*, 27-28 (Penguin, Harmondsworth, 1960).
10. Dibble, W. E. *J. hist. Astr.* **7**, 141-142 (1976).
11. Hawkins, G. S. *Stonehenge Decoded* (Souvenir, London, 1966).
12. Atkinson, R. J. C. *Antiquity* **31**, 219-233 (1957).
13. Atkinson, R. J. C. *Antiquity* **41**, 92-95 (1967).
14. Hoyle, F. *On Stonehenge*, 61 (Heinemann, London, 1977).

Determination of critical variables in a microbial predator-prey system by catastrophe theory

THE differential or difference equations that are used customarily to describe predator-prey interactions can give many insights into the behaviour of simple ecosystems, but only rarely can they be used to predict the population densities of real organisms with any accuracy. Thus even given long-term observations of population densities it is often difficult to deduce very much about the behaviour of the organisms. We present here an analysis in which we use population measurements to determine the variables that govern the response of an amoebal predator to its prey. Our results support the suggestion of Curds and Cockburn¹ that it is the prey-predator ratio rather than the prey density to which the amoebae are sensitive. We have arrived at our conclusions by applying catastrophe theory² in a way that we believe will be useful in many other contexts.

Dent *et al.*³ grew vegetative amoebae of the cellular slime mould, *Dictyostelium discoideum*, in continuous culture with a bacterial food source, *Escherichia coli* B/r. Figure 1 illustrates

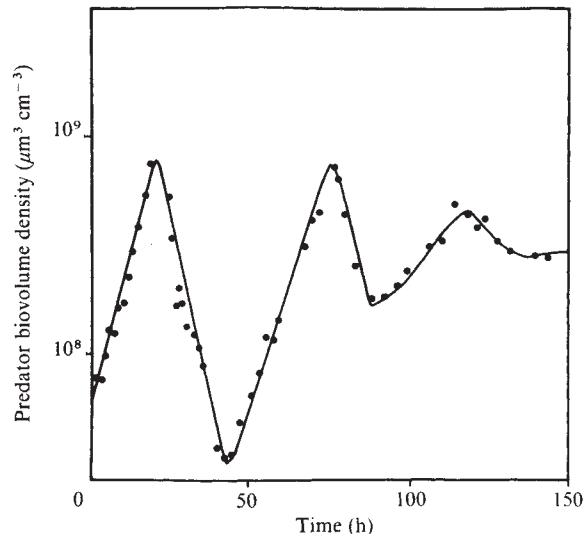


Fig. 1 Biovolume density of *D. discoideum*, data from ref. 3. The solid line is the fit obtained using the cusp catastrophe and is not significantly sensitive to the choice between the two conventions discussed in the text.

how during the first 150 h of the experiment the amoebae passed through alternating phases of exponential growth and decline. There was a very rapid change from one such phase to the next, and the specific growth rate, λ , remained constant throughout intervals of as much as a day, during which time the bacterial concentration measured in terms of biovolume (the product of the number density and the mean cell volume) could change by a factor of 100. The observation that the specific growth rate was shifting abruptly from one more or less constant value to another rather than moving smoothly in response to changes in prey density suggested that the system could profitably be studied in terms of catastrophe theory with λ as a state variable.

From Thom's list of seven elementary catastrophes we chose the cusp because it is the simplest which is consistent with the observation that in time the sudden changes in λ became smaller and eventually disappeared altogether. The cusp catastrophe requires two control variables, and we chose the culture age (a) and the bacterial biovolume density (H). With these choices a plot of prey biovolume as a function of time becomes the projection of the trajectory of the system into the control space (by system we mean the biochemical reactions which determine the behaviour of the amoebae, not the predator-prey system as a whole). By marking the points at which the abrupt changes in the state variable occurred, we were able to sketch the catastrophe set (Fig. 2). We could then relate a and H to the control variables of the canonical cusp catastrophe, and this in turn reduced the determination of λ as a function of a and H to a simple curve-fitting exercise. We then integrated λ to obtain the solid curve drawn in Fig. 1.

The curve clearly fits the data well, though we stress that the significant point is the qualitative picture, that is the nearly straight lines and the decreasing periods and slopes. The close agreement between the observed and predicted numerical values of the slopes is not evidence in favour of the model, because catastrophe theory allows us sufficient freedom to arrange this agreement providing that the qualitative behaviour is correct.

There is, however, a difficulty. Catastrophe theory explains sudden jumps in terms of shifts of the system from one stable equilibrium to another⁴. To predict where these will occur we must know not only how many stable steady states exist at a given point in the control space but also what determines which of two or more alternative steady states the system will be in. The former is deduced from the canonical form as given by Thom² and the latter is determined by specifying one of the 'conventions' of catastrophe theory. The two commonly used

conventions are the 'perfect delay' convention, in which we suppose that the system will remain in a local minimum until that minimum disappears, and the 'Maxwell' convention, in which we suppose that the system will always seek a global minimum of potential.

The catastrophe set indicated by the broken lines in Fig. 2 has the familiar cusp shape characteristic of the perfect delay convention, but the jumps occur as the trajectory enters the cusp, not as it leaves it. This implies that the system is moving away from one stable equilibrium and towards another at a higher potential. Such behaviour is not impossible to explain, but it involves a rather complicated mechanism as the system requires some separate means for choosing between steady states (P.T.S. in preparation).

We have found, however, that if we replace the control variable H by H/P , the ratio of prey to predator biovolume densities, we obtain the result shown in Fig. 3. All five jumps now lie on a single straight line, which we take as the catastrophe set of the cusp catastrophe with the Maxwell convention. This convention does not require us to postulate a separate switch, and so we deduce that the factor which determines the growth of the amoebae is more likely to be the ratio of biovolumes. This is contrary to what is assumed in most models of microbial predation but is supported by the result of Curds and Cockburn¹ who found that the feeding rate of *Tetrahymena pyriformis* was more closely correlated with H/P than with H .

We have a predator-prey system which we have not been able to model in detail and which has features inconsistent with the usual models. As a first step towards understanding the system, we want to know whether it is H or H/P to which the amoebae respond. Catastrophe theory tells us that if the critical variable is H/P , then a comparatively simple mechanism can account for the observed behaviour. If, however, the critical variable is the prey density H , then the mechanism will have to be much more complicated with, in effect, two separate switches to control a single response. We conclude that H/P is far more likely to be the correct choice.

The curves in Figs 2 and 3 were obtained by fitting cubic splines to the data of Dent *et al.*³ and we must ask how sensitive our results are to small variations in the shapes of these curves. It is clear that in Fig. 2 no minor change in the curves nor difference of an hour or two in the location of the peaks and troughs of amoebal biovolume can alter the configuration of the jumps. The catastrophe set will remain qualitatively the same and neither the Maxwell nor perfect delay conventions will be

Fig. 2 Biovolume density of *E. coli.*, with the catastrophe set indicated. The points at which the sudden changes in λ occurred are marked by solid circles. Data are based on ref. 3.

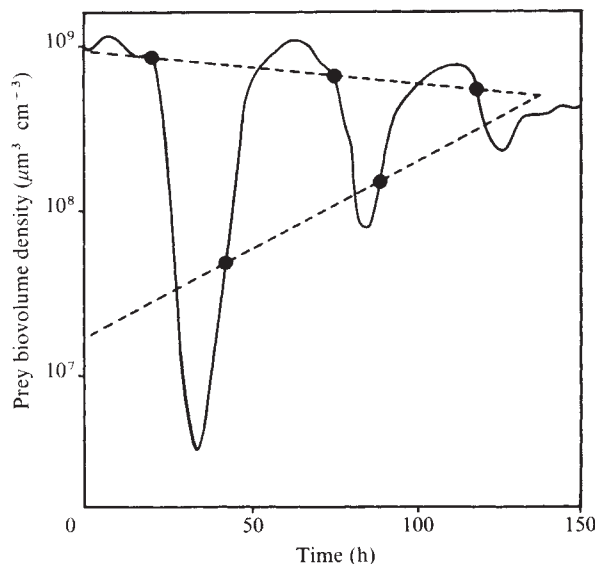
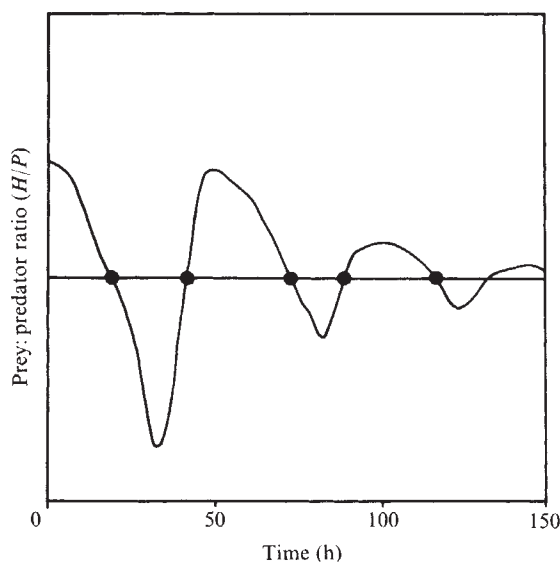


Fig. 3 The same as Fig. 2 but with prey-predator ratio as ordinate.

applicable. In the case of Fig. 3, however, we cannot rule out the possibility that the dynamic follows the perfect delay convention with a very narrow cusp. This would matter if we were trying to guess the actual mechanism, but it does not affect our argument concerning the relative simplicity of the possible mechanisms. Our conclusions thus do not depend on the precise shape of the curves.

If, as we believe, the rate of growth of amoebae is controlled by H/P , there must be some mechanism by which the amoebae can estimate this ratio. It has been suggested⁵ that folic acid, which is secreted by *E. coli* (and by soil bacteria as well) may serve as an attractant for food-seeking amoebae. If folic acid molecules are modified on contact with amoebal receptor sites, then the concentration will depend on the ratio of prey to predator densities. We are doing experiments to determine whether the folic acid concentration governs the behaviour of the amoebae.

We disagree with Zahler and Sussman⁶ that "catastrophe theory is a blind alley". By applying it we have discovered what seem to be the correct critical variables in our system. The theory made us concentrate on the amoebal specific growth rate, and ask how this could be affected. The need for two control variables led us to consider culture age as an explicit variable, even though most equations used in population dynamics are autonomous. This provided a link between the early stage of the experiment described here and the observation² that after about 200 h there is an apparently spontaneous change in the population densities of both organisms. Finally, catastrophe theory has provided strong evidence that it is the ratio of biovolume densities to which the amoebae respond, and this has suggested further experiments.

We expect that the way in which we have used catastrophe theory to decide between alternative choices of critical variables will find other applications in the study of complex systems. It does not require data as accurate as would be necessary to fit a model based on differential equations. It requires no definite hypothesis concerning the mechanism of the system and it leads to conclusions that are comparatively robust, for they do not stand or fall with one particular choice of mechanism or *ad hoc* equation. We do not claim that we are somehow obtaining more information from our analysis than was contained in the data—no mathematical technique can accomplish this. What catastrophe theory can do, however, is enable us to answer a limited number of questions about a system on the basis of much less information than would be needed for a complete description. Much of the power of the theory lies in the fact that this limited

number of questions so often turns out to include the most interesting ones.

We thank E. C. Zeeman, C. F. Thurston and J. L. Stirling for discussions. This research is supported in part by the NERC.

M. J. BAZIN
P. T. SAUNDERS

Department of Microbiology and
Department of Mathematics,
Queen Elizabeth College,
London W8, UK

Received 2 August 1977; accepted 19 June 1978.

1. Curds, C. R. & Cockburn, A. J. *gen. Microbiol.* **54**, 343–358 (1968).
2. Thom, R. *Stabilité Structurale et Morphogénèse* (Benjamin, Reading, Massachusetts, 1972).
3. Dent, V. E., Bazin, M. J. & Saunders, P. T. *Archs Microbiol.* **109**, 187–195 (1976).
4. Zeeman, E. C. *Sci. Am.* **234**, 65–83 (1976).
5. Pan, P., Hall, E. M. & Bonner, J. T. *Nature new Biol.* **237**, 181–182 (1972); *J. Bact.* **122**, 185–191 (1975).
6. Zahler, R. S. & Sussman, H. J. *Nature* **269**, 759–763 (1977).

Asynchronous sexual development determines the breeding system in field beans

THE breeding system of varieties of field bean (*Vicia faba* minor L.) has long been known to be intermediate between allogamy and autogamy, with approximately 1/3 of the seeds set as a result of cross-fertilisation and 2/3 by self-fertilisation¹. Cross-fertilisation is due to bees visiting open flowers. Most inbred lines show poor seed set unless their flowers are visited by bees or artificially manipulated (tripped) and are therefore called autosterile; a few lines which do not require tripping are called autofertile^{2,3}. Irrespective of variety, so far as is known, the field bean does not possess a self-incompatibility system for regulating fertilisation^{4–6}; also anther dehiscence always precedes flower opening by one or two days. Why then does the temporal advantage of self pollen over cross pollen not lead to obligate autogamy under field conditions? In an attempt to answer this question, it has been suggested that a spatial separation between discharged pollen and the stigma is normally present, but is abolished by bee visits which simultaneously bring about pollen–stigma contact and chance cross-pollination. According to this hypothesis, autosterile lines have a serious impediment to pollen–stigma contact, whereas, due to a different floral architecture, autofertile lines do not⁷, although intermediate floral structures also occur. Here we report results of a scanning electron microscopic (SEM) study of the stigmas of an autosterile line (T2) and an autofertile line (T51), which lead to an alternative explanation for autosterility and autofertility.

After freezing open flowers in vapourised liquid nitrogen at –110 °C the petals were removed without causing disturbance to the stigma region. Flowers of the autosterile line T2 examined

Fig. 1 Scanning electron micrographs of uncoated *Vicia faba* stigmas at 7.5 kV × 160. *a*, Line T2 after flower opening stage showing stigma surface covered with pollen grains: flower frozen in vapourised liquid nitrogen and dissected without stigma disturbance. *b*, Line T2 emasculated before anther dehiscence showing exudate on stigma surface 18 h after treatment (c.f. Fig. 2*d*, 2*e* for controls).

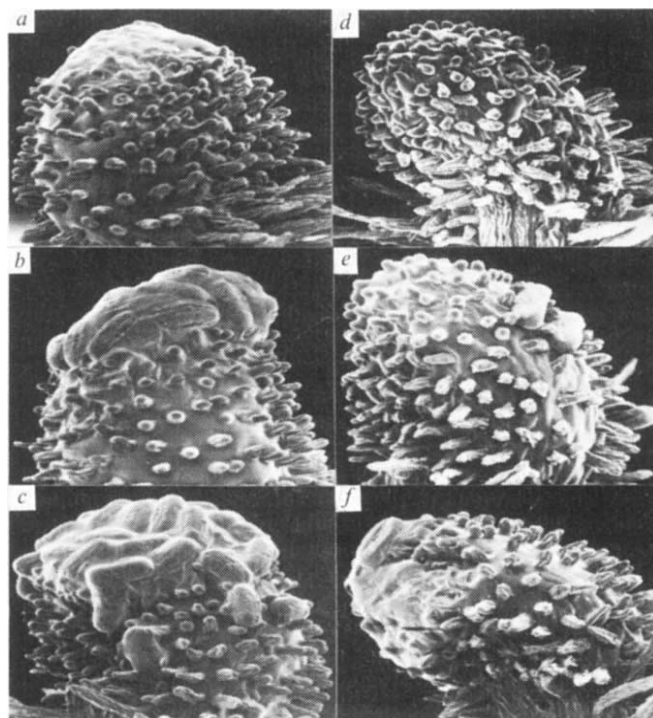
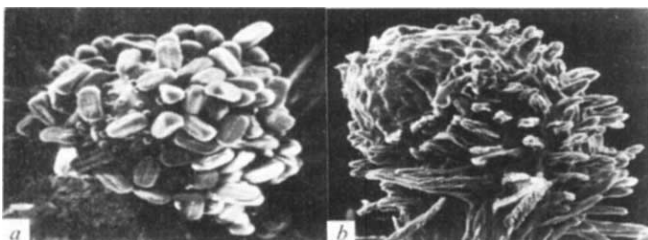


Fig. 2 Scanning electron micrographs of uncoated *Vicia faba* stigmas taken at 7.5 kV × 160, showing exudate appearance during flower development of line T51 (*a–c*) and line T2 (*d–f*). Flowers were taken from the 4th to 8th flowering node at comparable developmental stages: before anther dehiscence (*a*, *d*); after anther dehiscence but before flower opening (*b*, *e*); after flower opening (*c*, *f*).

in this way were found always to have stigmas covered with pollen grains (Fig. 1*a*), showing that autosterility is not due to a spatial separation of stigma and self pollen grains in the undisturbed flowers. Autofertile lines also showed abundant pollen. Stigma surfaces were then examined by SEM after ungerminated pollen grains had been removed from the stigma surfaces by briefly rinsing in distilled water. This investigation showed differences between lines T2 and T51 with regard to the amount and time of appearance of exudate on the receptive areas of the stigma tip (Fig. 2*a–f*). In the autofertile line T51, exudate was present before anther dehiscence (Fig. 2*a*). After dehiscence but before flower opening, self pollen was seen entrapped in an increased quantity of exudate (Fig. 2*b*). Even more exudate was apparent on the stigmas after the flowers opened (Fig. 2*c*). In the autosterile line T2, however, there was no exudate before anther dehiscence (Fig. 2*d*) or before flower opening (Fig. 2*e*), although small amounts of exudate were found on stigmas from open flowers (Fig. 2*f*).

In angiosperm species, including *Vicia faba*, in which a wet stigma is characteristic, it has been suggested that stigma receptivity is signalled by an accumulation of exudate on the stigma surface⁸. In autofertile line T51 and in some other lines (data not shown) where exudate appears and anthers dehiscence before flower opening, the probability of self-fertilisation might be expected to increase. This has been found to be so in at least one instance, where under field conditions⁹ an autofertile line showed much lower outcrossing than did corresponding autosterile lines.

In artificial field bean pollination, manual tripping of open flowers is a recommended practice for increasing seed set in autosterile lines¹⁰. Consequently, we examined the influence of tripping and emasculation on exudation and found, as expected, increased stigma exudate after both treatments (see Fig. 1*b*).

On the basis of these findings the so-called autofertile lines can be classified as protogynous and autosterile lines as protandrous. It seems that the timing of stigma receptivity, as manifested by exudation, acts as a critical control in field bean

pollination. This role of the stigma in actively regulating the breeding system may possibly apply to other bee-pollinated papilionaceous species.

We thank Gillian Toynbee-Clarke and Dr D. Bond (the Plant Breeding Institute, Cambridge) for seed material. C. P. acknowledges financial support from Deutsche Forschungsgemeinschaft and P. Gates is in receipt of a SRC postgraduate studentship.

C. PAUL
P. GATES
N. HARRIS
D. BOULTER

Department of Botany,
University of Durham,
Durham, UK

Received 30 May; accepted 21 July 1978.

1. Sirks, M. J. *Meded. LandbHooesch. Wageningen* **26**, 40 (1923).
2. Hayes, J. D. & Hanna, A. S. *Z. Pflanzenzüchtg.* **60**, 315–326 (1968).
3. Toynbee-Clarke, G. *J. agric. Sci., Camb.* **77**, 213–217 (1971).
4. Drayner, J. M. *J. agric. Sci., Camb.* **53**, 387–404 (1959).
5. Holden, J. H. W. & Bond, D. A. *Heredity* **15**, 175–192 (1960).
6. Hanna, A. S. & Lawes, D. A. *Ann. appl. Biol.* **59**, 289–295 (1967).
7. Kambal, A. E., Bond, D. A. & Toynbee-Clarke, G. *J. agric. Sci., Camb.* **87**, 519–526 (1976).
8. Heslop-Harrison, Y. & Shivanna, K. R. *Ann. Bot.* **41**, 1233–1258 (1977).
9. Kambal, A. E. *J. agric. Sci., Camb.* **72**, 131–138 (1969).
10. Toynbee-Clarke, G. *J. agric. Sci., Camb.* **82**, 531–534 (1974).

Does colour provide an input to human motion perception?

A PUZZLING feature of the visual system of primates is the occurrence of a multiplicity of separate 'maps' of the visual environment, each of which has its own independent set of afferent connections. Perhaps these maps represent an early stage in the analysis of visual information by the brain. Electrophysiological studies have shown that depth is analysed in V_2 (area 18) (ref. 1), that colour-coded cells occur mainly in V_4 and that motion in the visual field is analysed in the posterior bank of the superior temporal sulcus² (Fig. 1). We report here an experiment which suggests that colour and movement are handled separately in the visual system and that information defined by wavelength cannot be processed by the brain's motion-detecting mechanisms. This finding is consistent with Zeki's demonstration that "cells of the movement area are not concerned with colour"^{2,3}.

Stimuli used were based on Julesz' random-dot stereograms⁴ (Fig. 2). The members of the pair of random-dot arrays are identical except that all the dots of a central square-shaped region are horizontally shifted to an equal extent in one display in relation to the other. On monocular inspection, only random dots are seen for either display, but when fused binocularly, the central shifted region stands out in depth from the rest of the dots, as a square having a sharp border; this may be described as an illusory contour, as it is not defined by the (irregular) dots of the displaced region.

Lu and Fender⁵ found that when such a display is presented with coloured dots against a contrasting colour background, stereoscopic depth is lost when the dots and the background have equal luminance, although the dots are clearly seen by colour contrast at isoluminance.

The use of random-dot stereograms has not been limited to stereoscopic experiments. In the present experiment, random-dot displays are not presented to give stereoscopic depth; they are optically superimposed, and are exposed alternately at a rate which gives apparent motion of the central displaced region, which then appears as a square oscillating horizontally^{4,6}. The square has (very much as for stereoscopic presentation) a clearly defined illusory border. Neither the border nor the motion are visible without alternation of the pair of dot displays. The

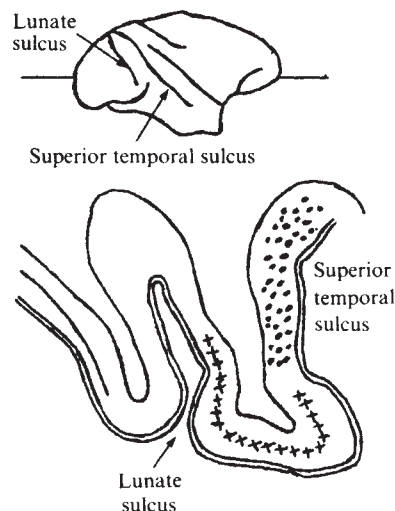


Fig. 1 The lower diagram shows a horizontal section of the brain of a rhesus monkey taken at the level indicated in the upper diagram to show regions of the prestriate cortex that are specialised for colour (marked by crosses) and motion (stippled) (adapted from Zeki²).

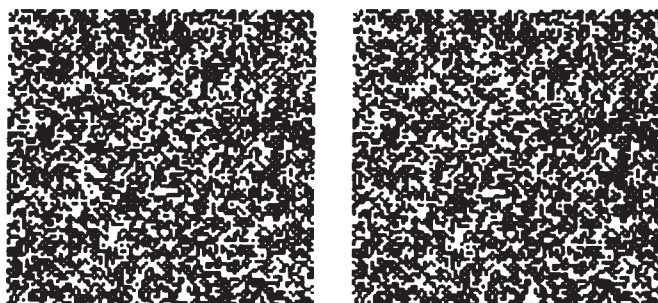
appearance of the border can be used as an independent and unambiguous criterion of motion perception⁷.

We used these patterns to study the role of colour as an input to human motion perception. Each dot in the original pattern (Fig. 2) is defined by a stepwise change in intensity (that is, black dots on a white background). Our strategy was to convert these intensity gradients into gradients of 'pure' wavelength; that is, we replaced all black-white borders with red-green borders using an isoluminance projection system⁸. The intensity of the green areas was then varied continuously over a wide range so that at least at some setting the intensities of red and green must have been precisely matched.

Isoluminance, with exact registration of the colour areas to prevent brightness contours, was obtained by projecting a special colour transparency with combined epi and dia projection through a single lens⁸. Two light sources were independently controlled for intensity, the epi source, which illuminates the upper coloured (red) surface of the opaque regions of the transparency, and the dia source, which projects by means of a colour filter (green) through the transparent regions. The transparency may be photographically produced silk screen prints; or 3M colour-key contact imaging material (red opaque negative), as used here. The format was 60 mm².

An identical pair of these projectors allowed a Julesz pair of transparencies to be superimposed, and projected alternately with a pair of electrically synchronised rotating sector-disk shutters. 'Key-stoning' distortion was avoided by placing one projector behind and the other in front of a translucent screen.

Fig. 2 A Julesz random-dot stereogram (see text for explanation). The two patterns were optically superimposed and exposed alternately to produce apparent motion.



The projected display was 37 cm² viewed from about 1 m, subtending 65°. The dot elements were 2 cm, subtending 3°, which was at least two orders above visual resolution at isoluminance. The screen luminance at isoluminance match was about 0.34 cd m⁻². Similar results were obtained at higher intensities (up to 34 cd m⁻²).

Four volunteers were used and two of them were unaware of the purpose of the experiment. Each subject began by deliberately introducing a large brightness difference between the red and green areas. At a frequency of about 100; 50; 100 (stimulus duration 100 ms; interstimulus interval or ISI 50 ms), a square with well-defined edges could be seen oscillating back and forth horizontally. Using the luminance control knob the intensity of the green dots was then varied continuously over a wide range. The subject was asked to record what he saw as he rotated the knob.

As the green background luminance was gradually reduced from a high value to near isoluminance, all subjects reported disappearance of the square. The square then reappeared when the luminance of the green was reduced to a level below the 'isoluminance' point. The disappearance of the square was quite sudden, and reported spontaneously by all four subjects.

Table 1 Critical range of luminance of green dots for each of four subjects

Subject	Luminance range (cd m ⁻²) over which square was invisible	Luminance range (cd m ⁻²) at which motion was invisible
1	4.42	6.84
2	4.32	6.84
3	4.73	7.68
4	3.49	7.68

Each entry represents the mean value of four readings.

Table 1 shows the 'critical range' of intensities over which the square disappeared for each of the four subjects. For all subjects: (1) the square disappeared slightly earlier than motion and was completely invisible over a narrow range of about 58% which was fairly constant for each subject, (2) the luminance setting for motion disappearance was remarkably constant among subjects. At this setting, three of the subjects reported that they could still see a slight 'shimmering' but none of them could detect any motion.

Having found the critical range, we varied other parameters to see if the square could be restored. Increasing the frequency of alternation did not make any difference. On lowering the frequency to about half the original rate, some of the dot elements could be seen moving, but the direction of motion was ambiguous and the edge defining the square remained invisible. Neither the square nor unambiguous motion could be restored at any frequency. We also tried doubling or halving the size of dot elements and/or the spatial displacement of the inner square, but these proved equally ineffective.

It is important to note that at the setting at which motion disappeared, the individual dots could be discerned just as clearly as when luminance contrast was present. Conversely, even gross blurring of the pattern (by defocusing the projectors) did not destroy the square as long as brightness differences were preserved.

In a control situation we used light red-dark red dot patterns instead of red-green. The intensity of one of the reds was then varied continuously. All four subjects reported that the disappearance of the square (and of motion) coincided almost exactly with the actual disappearance of the dots at isoluminance. The critical range for the disappearance of the square and of motion was thus much narrower than that measured using red-green dots (Table 1). Thus, colour cues seem to have, if anything, a slightly detrimental effect on the perception of coherent apparent motion.

The disappearance of motion at isoluminance is exactly analogous to Lu and Fender's⁵ discovery that stereopsis disappears at isoluminance in stationary random-dot stereograms. 'Classical' stereopsis, using monocularly visible line targets, apparently does not break down completely at isoluminance⁹, and we now find that this is also true of apparent motion. If two simple line targets are alternated instead of random-dot patterns, motion does not disappear at isoluminance.

As in the case of stereopsis⁹, random-dot targets may be processed differently from simple line targets. The former must involve cooperative interactions between large populations of cells tuned to either the same disparity^{4,10,11} or even (in the case of motion detectors) the same direction and velocity. Our findings may imply that at isoluminance cooperative facilitation is not possible between colour contrast-sensitive motion detectors. Alternatively, the relevant motion detectors may be insensitive to wavelength, which is consistent with physiological evidence³, and especially with the observation that monkey colour-opponent neurones give sustained responses whereas broad-band non-opponent neurones respond transiently¹²⁻¹⁴.

Regardless of our interpretation, it is clear from our findings that colour and motion are handled separately by the human visual system and that colour provides only a weak 'cue', at best, to movement perception.

We thank Dr O. J. Braddick for providing some of the Julesz stereograms used for making the special transparencies for isoluminance projection.

V. S. RAMACHANDRAN*
R. L. GREGORY

MRC Brain and Perception Laboratory,
Department of Anatomy,
The Medical School,
University of Bristol, Bristol, UK

Received 5 May; accepted 18 July 1978.

* Present address: Kenneth Craik Laboratory, Department of Experimental Psychology, University of Cambridge, Cambridge, UK.

- Hubel, D. H. & Wiesel, T. N. *Nature* **225**, 41-42 (1970).
- Zeki, S. M. *Proc. R. Soc. Lond. B* **197**, 195-223 (1977).
- Zeki, S. M. *J. Physiol., Lond.* (1974).
- Julesz, B. *Foundations of Cyclopean Perception* (University of Chicago Press, 1971).
- Lu, C. & Fender, D. H. *Invest. Ophthalmol.* **11**, 482-489 (1972).
- Anstis, S. M. *Vision Res.* **10**, 1411-1430 (1970).
- Braddick, O. J. *Vision Res.* **13**, 355-369 (1973).
- Gregory, R. L. *Perception* **6**, 113-119 (1977).
- Ramachandran, V. S., Rao, V. M. & Vidyasagar, T. R. *Nature* **242**, 412-414 (1973).
- Nelson, J. I. *J. theor. Biol.* **49**, 1-88 (1975).
- Marr, D. & Poggio, T. *Science* **194**, 283-287 (1976).
- Gouras, P. J. *Physiol., Lond.* **199**, 533-547 (1968); **204**, 407-419 (1969).
- De Monasterio, F. M., Gouras, P. & Tolhurst, D. J. *Vision Res.* **16**, 674-678 (1976).
- Dreher, B., Fukada, Y. & Rodieck, W. J. *Physiol., Lond.* **258**, 433-452 (1976).

Suppression of oestrogen-induced LH surges by social subordination in talapoin monkeys

IT is axiomatic that the social structure of a group of primates has a pervasive effect on each individual's behaviour. Thus, many monkeys form dominance hierarchies, based on the successful outcome of aggressive interactions, and the more subordinate animals show characteristic behavioural traits^{1,2}. In particular, the social hierarchy has marked effects on reproductive behaviour, especially in talapoin monkeys, where sexual interactions tend to be the prerogative of the dominant males³. Subordinate male rhesus monkeys have limited access to sexually attractive females and produce fewer offspring⁴, and in the few species that have been carefully examined, subordinate females, although apparently receiving sexual attention, are, nevertheless, less fertile than would be expected⁵⁻⁷. Recently, it has become clear that the dominance hierarchy affects not only behaviour but also hormone levels. Testosterone is higher in dominant males^{8,9}, cortisol and prolactin are raised in more subordinate animals¹⁰. We report here evidence that subordination can prevent the luteinising hormone (LH) surge which

Table 1 Sexual and aggressive behaviour, plasma cortisol and prolactin in female talapoin

		Dominant female	Subordinate female	<i>z</i>	<i>P</i>
Total aggression received from males	Attacks	45	110	2.099	<0.02
	Threats	11	15	0.54	NS
	Displaces	8	61	3.79	<0.001
Total aggression received from females	Attacks	2	55	3.52	<0.0002
	Threats	0	11	2.28	<0.01
	Displaces	0	216	8.79	<0.0001
Sexual behaviour received	Mounts	339	64	7.31	<0.0001
	Ejaculations	139	6	7.52	<0.0001
Mean plasma prolactin $\pm$ s.d. (ng ml ⁻¹ , human standard)		61.4 $\pm$ 43.0	125 $\pm$ 55.4	<i>t</i> 8.22	<0.001
Mean plasma cortisol (μ g ml ⁻¹ $\pm$ s.d.)		176.5 $\pm$ 43.1	207.6 $\pm$ 76.6	2.4	<0.02

Behaviour was based on 1,000 min of observations during five separate two-week periods when females received oestrogen. Plasma samples were collected over a 10-month period during which females received their oestrogen implants. Attack: the aggressor chases its opponent and attempts to grab and bite it. Threat: eyebrows are raised, eyes open wide and the mouth open to varying degrees. Displace: such aggression often occurs when the subordinate monkey is occupying a food hopper or sitting with another monkey. Without necessarily making contact the dominant monkey may cause the subordinate to move away and then occupy the same position. (*z*, Mann-Whitney; *t*, Student's *t*-test). NS, not significant.

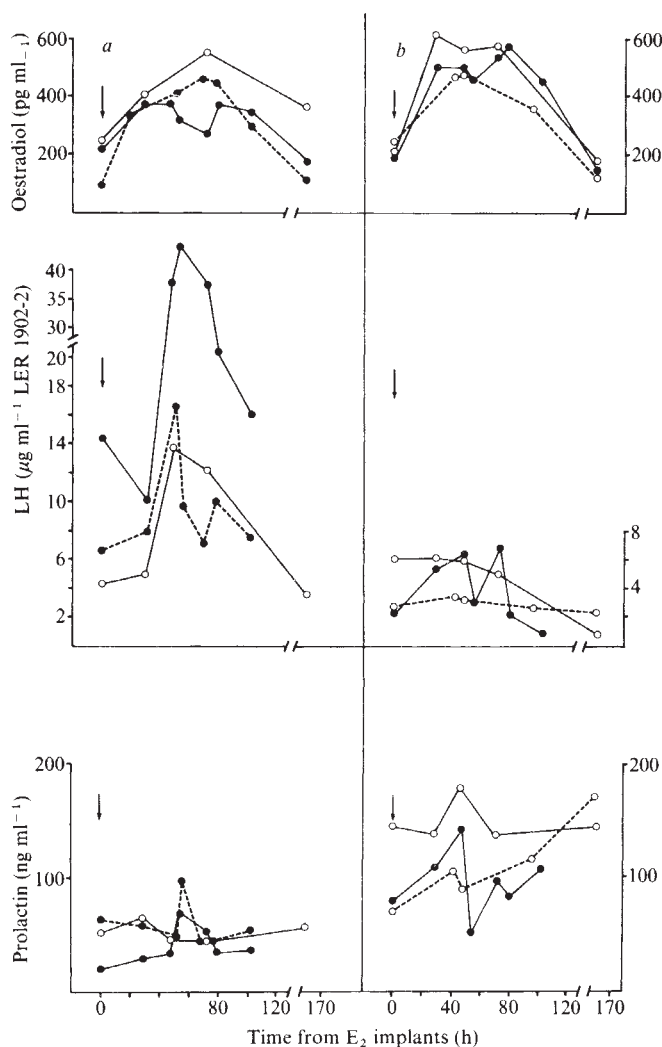
is normally induced by giving oestrogen to female monkeys, thus offering an explanation for reduced fertility and suggesting that this can be induced by behavioural interactions.

A socially living group of four male and four female talapoin was studied in a large cage (11' $\times$ 8' $\times$ 5') in which they had lived for about one year before the experiments described here. Males and females were denied access to each other apart from for two periods of 50 min d⁻¹, when sexual, aggressive and social behaviour were scored quantitatively from behind a one-way vision screen as described elsewhere¹¹.

The males were intact; females were ovariectomised and were implanted with a single oestrogen-filled capsule (1.5 cm $\times$ 3.2 mm $\times$ 1.0 mm) producing plasma levels of 100–150 pg ml⁻¹, approximating those occurring in intact females during the follicular phase of the cycle. These were removed and replaced at intervals to ensure maximal sexual skin swelling and female attractiveness. The administration of an oestrogen surge to the hypothalamo-pituitary axis is a well recognised method of inducing discharge of LH in primates¹². An oestrogen surge was produced in the female talapoin by inserting two additional oestrogen-filled capsules and plasma samples were then taken at intervals over the following 5 d. Plasma oestradiol¹³, LH¹⁴ and prolactin¹⁵ were measured by radioimmunoassay and the animals' behaviour continued to be scored twice daily.

Before oestrogen surges were given, the group had been observed for about one year and plasma samples taken twice weekly. During this time, both dominant and subordinate females received the same dose of oestrogen. Both engaged in sexual interactions, although this was more frequently observed in the case of the dominant female (Table 1). Aggressive behaviour was also different, the subordinate female receiving much more than the dominant. Plasma cortisol and prolactin levels were raised in the subordinate compared to the dominant female during the time they lived together (Table 1). The dominant female when given an oestrogen surge on three occasions (each separated by approximately 21 d) always produced an LH surge about 50–70 h later; in no case was this observed in the subordinate female even though oestrogen levels were comparable to those of the dominant animal (Fig. 1). During treatment with additional oestrogen, the behavioural differences between the two females were again observed, the dominant female receiving more ejaculations (18 mounts and 15 ejaculations in 10 h observation) than the subordinate (14 mounts, 0 ejaculations in 10 h observation). No aggressive interaction was directed towards her although a considerable amount was received by the subordinate (10 attacks, 4 threats,

Fig. 1 Plasma LH and prolactin levels in dominant (*a*) and subordinate (*b*) female talapoin monkeys when given extra oestradiol (E₂) implants (indicated by arrows) over a five-day period. The subordinate female has significantly higher prolactin levels and unlike the dominant female does not respond with an LH surge.



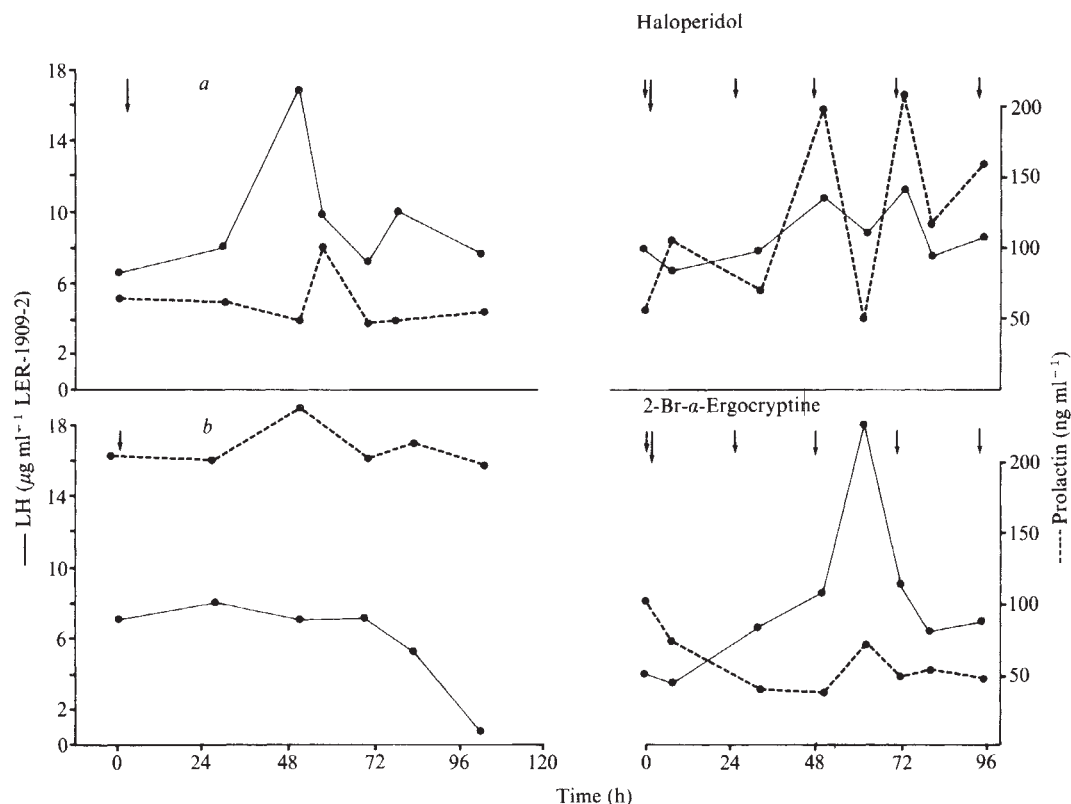


Fig. 2 Effects of suppressing or elevating prolactin on the oestrogen (indicated by large arrow) induced LH surge. The subordinate female (b) receives bromocryptine treatment by daily injection (indicated by small arrows) which lowers plasma prolactin and permits an LH surge. The LH surge is blocked when plasma prolactin is elevated by haloperidol treatment in the dominant female (a).

14 displaces). Plasma prolactin continued to be significantly higher in the subordinate female (mean $115.7 \pm 35.3 \text{ ng ml}^{-1}$) than the dominant female ($50.8 \pm 17.0 \text{ ng ml}^{-1}$), $t = 7.15$, $P < 0.0001$) during the oestrogen challenge (Fig. 1). The subordinate female was subsequently treated with a dopamine agonist, bromocryptine, a procedure which effectively lowers prolactin in humans and can restore ovulation^{16,17}. This resulted in the reduction of plasma prolactin levels, and when challenged again with two oestrogen-filled capsules during bromocryptine treatment, an LH surge occurred which resembled that formerly seen in the dominant female. Conversely, the prolactin levels in the dominant female were raised by giving her the dopamine receptor blocker, haloperidol. This both raised her prolactin level and rendered the oestrogen challenge ineffective, no additional LH being released (Fig. 2). (It is also well recognised in humans that high prolactin levels are associated with amenorrhoea and failure to ovulate¹⁶.)

As it was possible that the raised prolactin levels in the subordinate female might be due to factors other than her social environment, she was removed from the group and housed singly. Eight weeks later, her prolactin levels had fallen to those comparable with the more dominant animals, and the oestrogen challenge now induced an LH surge without the need for bromocryptine treatment.

Our results show that behavioural interactions can so alter the female primate's neuroendocrine system as to render it unlikely that she could conceive. Thus, reduced fertility observed in certain subordinate female primates in the wild, and that observed in humans during stressful situations¹⁸, could be related to the hormonal changes we describe here. We cannot be sure that the effect of the drugs on LH secretion was due to the consequent changes in prolactin, for dopamine is also implicated directly in the release of LH¹⁹. It is certainly possible that high levels of prolactin, however produced, are inhibitory to ovulation and hence fertility. Furthermore, we can speculate that the syndrome of hyperprolactinaemia in women could be induced or aggravated by behavioural mechanisms. Of course, these would not necessarily involve overt aggression as described in the talapoin monkeys, but whatever factors threaten the social stability of the individuals concerned, and the way in which they

react to this, would provide a more likely parallel for our own species. While not wishing to imply that all subordinate monkeys are necessarily infertile, such a mechanism, if it is physiologically active, may also be a way whereby subordinate females would contribute less to the breeding of a group than might be expected by simply observing their overt sexual behaviour.

This work was supported by an MRC programme grant to Joe Herbert whose help has been invaluable throughout. We thank Raith Overhill for drawing the figures, Debbie Steer for typing the manuscript, Drs Friesen (PRL antiserum) and Niswender (LH antiserum) NIAMDD (LH and PRL standards), MRC (h-PRL for iodination) for materials for radioimmunoassay, Sandoz for bromocryptine and Searle for haloperidol.

LESLEY A. BOWMAN
SUSAN R. DILLEY
E. B. KEVERNE*

Department of Anatomy,
University of Cambridge,
Cambridge, UK

Received 14 June; accepted 25 July 1978.

* To whom reprint requests should be addressed.

- Richards, S. M. *Anim. Behav.* **22**, 914-930 (1974).
- Deag, J. M. *Anim. Behav.* **25**, 465-474 (1977).
- Dixon, A. F. & Herbert, J. *Horm. Behav.* **8**, 141-154 (1977).
- Duvall, S. W., Bernstein, I. S. & Gordon, T. P. *J. Reprod. Fert.* **47**, 25-31 (1977).
- Drickhamer, L. C. *Folia Primatol.* **21**, 61-80 (1974).
- Dunbar, R. I. M. & Dunbar, E. R. *Nature* **266**, 351-352 (1977).
- Epplé, G. *Folia Primatol.* **7**, 37-65 (1967).
- Rose, R. M., Bernstein, I. S. & Holaday, J. Q. *Nature* **231**, 366-368 (1971).
- Keverne, E. B., Meller, R. E. & Martinez-Arias, A. *Rec. Adv. Primatol.* **1**, 533-548 (1978).
- Keverne, E. B. in *Sex, Hormones and Behaviour* (Ciba Symp., in the press).
- Dixon, A. J., Scruton, D. M. & Herbert, J. *J. zool. (Res.)* **176**, 177-210 (1975).
- Knobil, E. *Rec. Prog. Horm. Res.* **30**, 1-46 (1974).
- Exley, D., Johnson, M. W. & Dean, P. D. *G. Steroids* **18**, 605-620 (1971).
- Niswender, G. D., Reichert, L. E., Midgley, A. R. & Nalbandov, A. V. *Endocrinology* **34**, 1166-1173 (1969).
- McNeilly, A. S. *Proc. R. Soc. Med.* **66**, 863-864 (1973).
- Thorner, M. O., McNeilly, A. S., Hagen, C. & Besser, G. M. *Br. med. J.*, 419-422 (1974).
- Besser, G. M., Parke, L. Edwards, C. R. W., Forsyth, I. A. & McNeilly, A. S. *Br. med. J.* **669**-672 (1972).
- Zacur, H. A., Chapanis, N. O., Lake, C. R., Ziegler, M. & Tyson, J. E. *Am. J. Obstet Gynec.* **125**, 859-862 (1976).
- Kordon, C. & Ramirez, V. D. in *Anatomical Neuroendocrinology* (eds Stumpf, W. E. & Grant, L.) 409-419 (Karger-Basel, 1975).

Lysosomal and mitochondrial heat labile enzymes in ageing human fibroblasts

ONE of the characteristic changes associated with the ageing of human fibroblasts in cell culture is the presence of increasing amounts of altered molecules¹⁻³. Such alterations have been detected for at least three cytoplasmic enzymes: glucose-6-phosphate dehydrogenase (G6PD), 6-phosphogluconate dehydrogenase and hypoxanthine-guanine-phosphoribosyl-transferase. For each enzyme, the proportion of molecules which are modified reaches as much as 25% in cells of the last generations. The origin of these alterations is still only an hypothesis. Some authors¹, in agreement with Orgel's error catastrophe theory⁴, suggest that they are the result of a faulty synthesising machinery; others⁵⁻⁷, however, suggest that they are due to a post-translational effect. We present here a study on the heat lability of some lysosomal and mitochondrial enzymes to test whether the alterations also affect proteins present in these subcellular entities. We compare the results with the predictions of the above hypotheses, and suggest that they probably are best explained in terms of a post-translational modification.

Experiments were carried out on normal human fibroblasts (WI-38). Using Hayflick's method⁸, these cells were cultured and the population was split 1:4 every week. Cells from confluent cultures were collected, homogenised and heated to high temperatures so as to control the denaturation rates of the enzymes (Fig. 1 and Table 1). Five enzymes were tested: *N*-

acetyl- β -D-glucosaminidase (NA β Glu), α -D-glucosidase (α Glu) and *N*-acetyl- α -D-galactosaminidase (NA α Gal), all located in the lysosomes, and sulphite cytochrome *c* reductase and G6PD, located in the mitochondria and cytoplasm, respectively. For NA β Glu, only isoenzyme B was considered. Isoenzyme A, which had first been inactivated by incubating the homogenate for 60 min at 37 °C in an imidazole-HCl buffer at pH 7 (ref. 11), was responsible for 10% of the enzyme activity in both young and old cells. The protein concentration of the homogenate was adjusted to obtain accurate assays on the enzymes (Table 1). The slight variations in the protein content of the homogenate did not affect the denaturation rates of the enzymes but we added albumin to serve as a protective agent when high dilutions of this homogenate were required. On the other hand, temperature and pH controls were crucial in such experiments. Two cell populations from the same line were compared: those from passages 22-25 and those from passages 42-45, which had completed 46-53% and 89-96% of their maximum lifespan, respectively. The maximum lifespan was observed on cells of the same lineage.

The experimental points presented in Fig. 1 were subjected to a factorial two-way analysis of the variance and showed a non-significant time-age interaction except for a significant interaction with G6PD ($P < 0.01$). When this enzyme was tested in the old cells, the deviation of the curve from linearity was significant ($P < 0.05$) and the experimental values fitted best by superposing two decreasing exponentials; the first one, with the steeper slope, represented the denaturation rate of the heat-labile G6PD, and the second one, that of the unaltered enzyme. The latter was responsible for 77% of the enzyme activity. Thus, in the analysis made in our laboratory, only the G6PD was altered in the old cells. The experiments were

Fig. 1 Inactivation curves of enzymes from young and old WI-38 fibroblasts. Confluent cultures were rinsed with phosphate buffered saline (PBS) and the cells were collected with a rubber policeman. They were then washed twice with 0.25 M sucrose solution and homogenised by 20 strokes of the tight pestle of a Dounce homogeniser (Kontes). For the assay of sulphite cytochrome *c* reductase, the cells were disrupted with an ultraturax. The homogenates were heated in the incubation medium to the temperatures critical for the denaturation of the enzymes. At given times, samples were removed and tested for their residual enzyme activity. The experimental conditions are summarised in Table 1. Enzyme assays were introduced following methods already published but with slight modifications, that is, NA β Glu, α Glu and NA α Gal were tested with umbelliferyl derivatives⁹ at pH 3.5, 4.5 and 4, respectively; sulphite cytochrome *c* reductase as originally described at pH 8.5 (ref. 10); the preparation of the extract and the assay of the G6PD were carried out following the method of Holliday and Tarrant¹. Results were expressed

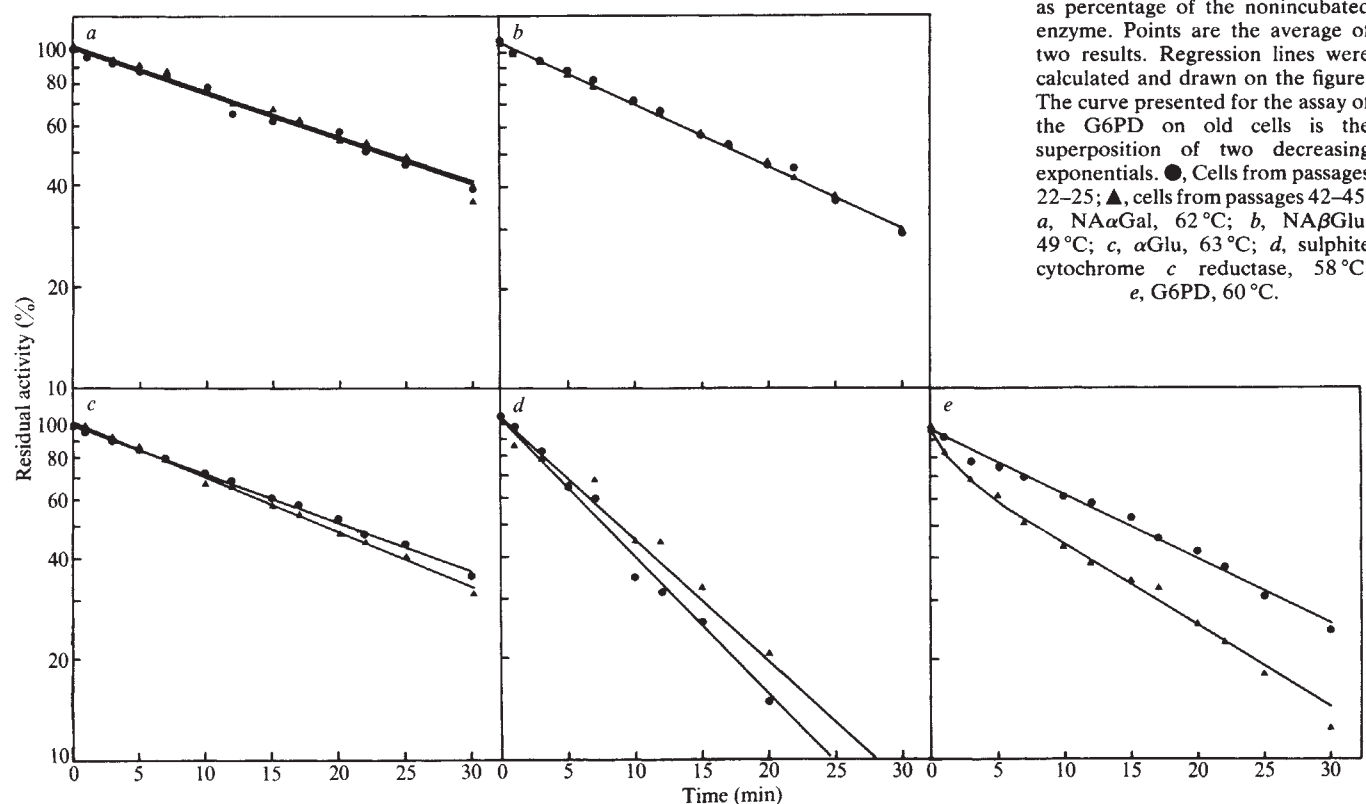


Table 1 Experimental conditions for the inactivation curves

Enzyme tested	Range of the protein concentration of the homogenate (mg ml ⁻¹)	Buffer	pH	Other components	Temperature of incubation (°C)
N-acetyl- α -D-galactosaminidase	1–1.5	5 mM acetate	4	Albumin 1 mg ml ⁻¹	62
N-acetyl- β -D-glucosaminidase	0.003–0.005	50 mM citrate	3.5	Albumin 1 mg ml ⁻¹	49
α -D-glucosidase	0.1–0.15	10 mM acetate	4.5	Albumin 1 mg ml ⁻¹	63
Sulphite cytochrome c reductase	5–8	3 mM imidazole-HCl	8.5	Sucrose 0.25 M	58
Glucose-6-phosphate dehydrogenase	1.5–2	50 mM Tris-HCl	8	Mercaptoethanol 1 mM EDTA 1 mM ϵ -amino-n-caproic acid 1 mM NADP 1 mM	60

repeated at least twice, as described above, and once again at a lower temperature (2 °C below that indicated); within a range of 2–5% no alteration could be detected for the lysosomal and mitochondrial enzymes.

Holliday and Tarrant¹ showed that some cytoplasmic enzymes are modified during the ageing of human MRC 5 fibroblasts in culture. We confirmed these results with WI-38 cells and found that the percentage of heat-labile G6PD observed in old populations is similar in both cell strains. To explain the appearance of altered enzymes in old cells, Orgel's error catastrophe theory of ageing¹² was proposed¹. This theory suggests that all the enzyme molecules would be modified, as they are synthesised by faulty machinery. However, we show that at least four enzymes are not significantly modified during the ageing process. One explanation for the fact that modifications would not be detectable in these molecules is that the altered enzymes would be completely inactive. This eventuality is rather improbable, for a misincorporation of amino acids produces considerable heterogeneity in the loss of enzyme activity¹³. The hypothesis of a post-translational modification explains more easily how alterations occur in only some enzymes. Several results have been published supporting this explanation^{14–18}, but discussion on the subject remains open^{19–21}. The presence of unaltered enzymes in lysosomes and mitochondria, apart from the modified cytoplasmic enzymes, also suggests that the alteration could be specific to certain cellular compartments. However, the analysis of many more enzymes would be necessary before we could generalise this hypothesis. Note also that not all the cytoplasmic enzymes are modified, for phosphoglucose isomerase was found to be unaltered in old cell cultures²². In conclusion, the fact that four enzymes located in lysosomes and mitochondria are not modified during the ageing of cell cultures suggests that the error catastrophe theory is not a satisfactory explanation for the presence of such alterations in some cytoplasmic enzymes.

We thank Mrs A. Genette, E. Feytmans and E. Depiereux for the statistical analysis of the results and Mrs A. Houbion for assistance. The work was supported by the "Belgian Fondation Universitaire pour l'étude des processus du vieillissement". The WI-38 cells were obtained from the American type culture collection.

A. HOUBEN
J. REMACLE*

Unité de Biochimie Cellulaire
et Laboratoire de Chimie Physiologique,
Facultés Universitaires, Namur, Belgium

Received 23 February; accepted 12 June 1978.

* To whom reprint requests should be addressed.

- Holliday, R. & Tarrant, G. M. *Nature* **238**, 26–30 (1972).
- Lewis, C. M. & Tarrant, G. M. *Nature* **239**, 316–318 (1972).
- Goldstein, S. & Moerman, E. J. *Nature* **255**, 159 (1975).
- Orgel, L. E. *Proc. natn. Acad. Sci. U.S.A.* **49**, 517–521 (1963).
- Gershon, D. & Gershon, H. *Gerontology* **22**, 212–219 (1976).

- Baird, M. B., Samis, H. V., Massie, H. R. & Zimmerman, J. A. *Gerontologia* **21**, 57–63 (1975).
- Dreyfus, J. C. *et al. Gerontology* **23**, 211–218 (1977).
- Hayflick, L. *Expl Cell Res.* **37**, 614–636 (1965).
- Van Hoof, thesis, Univ. Louvain 285 (1975).
- Wattiaux-de Coninck, S. & Wattiaux, R. *Eur. J. Biochem.* **19**, 552–556 (1971).
- Robinson, D. & Stirling, G. *Biochem. J.* **107**, 321–327 (1968).
- Orgel, L. E. *Nature* **243**, 441–445, (1973).
- Langridge, J. *Molec. gen. Genet.* **103**, 116 (1968).
- Pendergrass, W. R., Martin, G. M. & Bornstein, P. *J. cell. Physiol.* **87**, 3–14 (1976).
- Danot, M., Gerson, H. & Gershon, D. *Mech. Ageing Dev.* **4**, 289–299 (1975).
- Steinhagen-Thiessen, E. & Hilz, H. *Mech. Ageing Dev.* **5**, 447–457 (1976).
- Kahn, A. *et al. Gerontology* **23**, 174–184 (1977).
- Duncan, M. R., Dell'Orco, R. T. & Cuthrie, P. L. *J. cell. Physiol.* **93**, 49–56 (1978).
- Baird, M. B., Samis, H. V., Massie, H. R. & Zimmerman, J. A. *Gerontologia* **21**, 57–63 (1975).
- Holliday, R. *Gerontologia* **21**, 64–68 (1975).
- Dreyfus, J.-C., Robinson, H., Schapira, F., Weber, A., Marie, J. & Kahn, A. *Gerontology* **23**, 211–218 (1977).
- Shakespeare, V. & Buchaman, J. H., *J. cell. Physiol.* **94**, 105–116 (1978).

Initiation of chemical carcinogenesis requires cell proliferation

CELL proliferation has often been implicated in the development of cancer with chemicals^{1–3}. Supportive evidence for this is the observation that several carcinogens that normally do not induce liver cancer in intact adult animals, especially with a single dose, become carcinogenic if administered as a single injection after partial hepatectomy (PH)¹. In these conditions, it is thought that PH may act during initiation, presumably by fixation of some carcinogen-induced DNA damage through replication of the altered DNA. However, because carcinogenesis is a multi-step process, often involving the appearance of several new cell populations between the initial target cells and the ultimate cancer^{4,5}, it is possible that the regenerative response of liver following PH could have a major effect on one or more steps subsequent to initiation. Thus, the use of a very late end-point (cancer) makes it difficult, if not impossible, to relate with any accuracy some very early event to any specific step in the carcinogenic process. Solt and Farber⁶ have recently developed a new approach to the sequential analysis of carcinogenesis *in vivo* that delineates the first few steps in the process and is a quantitative assay for initiation of liver cancer. We have used this model to investigate whether cell replication exerts its first effect on initiation or on some later step in the carcinogenic process. And if initiation is at least one site of action, when in the regenerative cell cycle is a carcinogen most effective and what biochemical events might be involved at this phase of the cell cycle?

The assay⁶ is based on the hypothesis that an early, if not the first, cell is resistant to the inhibitory effect of carcinogens on cell proliferation. It consists of the measurement of the number of foci of carcinogen-altered hepatocytes that are capable of responding to a stimulus for cell proliferation in the presence of a strong inhibitor of normal cell proliferation (for example, low dietary level of 2-acetylaminofluorene, 2-AAF). The combined dietary 2-AAF plus PH creates a strong selection pressure for

the growth of carcinogen-induced putative initiated cells, but does not by itself induce such changes in the experimental conditions used^{6,7}. The foci of altered cells can readily be identified either grossly or with the histochemical use of γ -glutamyl transpeptidase (γ -GT) activity, an enzyme that is a marker for pre-neoplastic liver lesions in these conditions⁷⁻¹⁰. In addition, with a single dose of at least one chemical carcinogen, diethylnitrosamine (DEN), as the initiating carcinogen and with this model, hepatocellular carcinoma develops in the majority of rats within 8 months⁷. The cancers are seen to arise within hyperplastic nodules which in turn are derived from foci of altered resistant hepatocytes⁷. For example, with a large initiating dose of DEN, such as 200 mg per kg, approximately 10^3 foci appear in a 5-g liver during selection, of which about 10^2 develop into persistent hyperplastic nodules. One to five hepatocellular carcinomas develop from such nodules.

For this study, a potent carcinogen for many tissues or organs and one not requiring metabolic activation, *N*-methyl-*N*-nitrosourea (MNU), was selected. This compound is not a carcinogen for intact liver despite its extensive methylation of DNA and other liver macromolecules¹¹, but becomes one when administered after partial hepatectomy¹². MNU was administered intraperitoneally (i.p.) to male Fischer rats (F-344). Three hours later, a time at which no further MNU could be detected in the blood or in the liver (E.C. and D.S.R.S., unpublished), the animals were subjected to either 67% partial hepatectomy or sham operation (SO). After a recovery period of 2 weeks, all the rats were exposed to the selection procedure favouring the growth of the putative initiated hepatocytes over the surrounding hepatocytes. This consisted of dietary 2-AAF plus the liver mitogen, α -hexachlorocyclohexane (α -HCH)¹³, as shown in Fig. 1.

The results presented in Table 1 show that foci of resistant phenotypically altered hepatocytes, positive for γ -GT, were seen in rats which received MNU and which were partially hepatectomised. Very few foci were seen in the animals receiving MNU + SO, vehicle + PH or no MNU. A similar pattern of results was obtained using male Wistar rats. Also, when CCl_4 replaced the α -HCH as a stimulus for cell proliferation, significant numbers of γ -GT-positive foci were again seen only in rats receiving MNU plus PH. Recent results in experiments designed to determine the time relationships between the administration of MNU and PH for optimal initiation indicate that MNU is a much more efficient inducer of putative initiated cells when administered 18 h after operation (129 foci cm^{-2}). The degree of initiation falls off rapidly with administration of MNU before 16 h or after 24 h following partial hepatectomy.

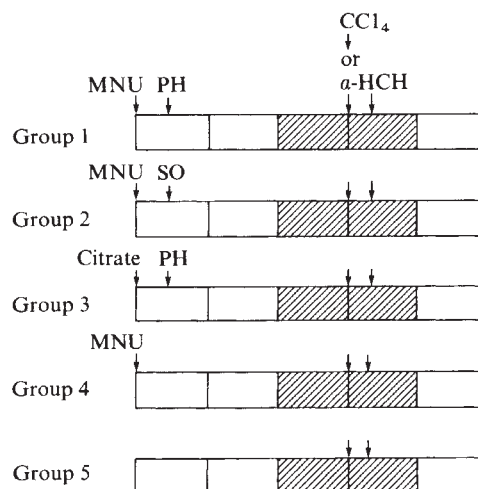


Fig. 1 Adult male Fischer 344 rats (Canadian Breeding Farms and Laboratories) were given a single dose of *N*-methyl-*N*-nitrosourea (60 mg per kg, i.p. in 10 mM citrate buffer pH 6.0). After 3 h they were either submitted to partial hepatectomy (PH) or sham operated (SO). Control groups received MNU and no operation, citrate buffer and PH, or citrate buffer and SO. Two weeks later, all the animals were placed on a diet containing 0.02% 2-AAF in order to create a 'mito-inhibitory' effect, and 1 week after starting the 2-AAF diet they were subjected to a stimulus for liver cell proliferation by force-feeding α -HCH, at a dose of 150 mg per kg dissolved in corn oil, on days 7 and 9 after commencing the 2-AAF diet, or a single dose of CCl_4 (0.25 ml per 100 g in corn oil). After a total period of 14 days on 2-AAF diet, the rats were placed on the stock diet devoid of AAF for 7 days. At the end of this period, they were killed and liver samples were processed for histological and histochemical studies. To monitor the effectiveness of the selection procedure, rats previously initiated with diethylnitrosamine (DEN: 100 mg per kg, i.p.) were also included at the start of the 2-AAF diet. Open bar, 1 week basal diet; hatched bar, 1 week basal diet + 0.02% AAF.

These data clearly indicate that initiation of carcinogenesis with MNU is intimately related to some change in the liver that occurs after removal of two-thirds of the liver.

We have not yet established that the administration of MNU in partially hepatectomised rats in the conditions used in these experiments leads to liver cancer. However, there is considerable circumstantial evidence in support of this. For example, (1) MNU in similar conditions has been found to be carcinogenic for the rat liver in two laboratories^{12,14}; (2) the foci induced by MNU, like those induced by DEN and other liver carcinogens, are positive for γ -GT, a marker enzyme for pre-neoplastic and neoplastic hepatocytes⁸; and (3) the foci are indistinguishable in their appearance and early behaviour from those induced by DEN and the latter have been shown to be precursors for liver cancer⁷.

Although the mechanism by which cell proliferation stimulates the induction of putative pre-neoplastic hepatocytes is not known, replication of carcinogen-damaged DNA before repair offers an attractive mechanism by which the carcinogen-induced DNA damage may be fixed and may thus increase the chances of altered nucleotide sequences in the newly made DNA. In support of this hypothesis, it has been demonstrated that liver DNA containing carcinogen-induced lesions such as those induced by MNU (D.S.R.S., S. Rajalakshmi and P. M. Rao, unpublished observations), dimethylnitrosamine^{15,16} and *N*-hydroxy-2-acetylaminofluorene¹⁷ replicates *in vivo*. Such a phenomenon is not lethal for the cell, since the newly made DNA on the damaged templates is stable *in vivo* for at least one month. Interestingly, several liver carcinogens are necrogenic and induce compensatory liver cell proliferation; some of these induce resistant foci without the need for an exogenous mitogen such as PH (refs 6, 7).

Table 1 Effect of partial hepatectomy on the induction of γ -GT-positive foci in the liver by MNU

Treatment	α -HCH		CCl_4
	Fischer	Wistar	Fischer
MNU + PH	$7.4 \pm 2.5(4)$	12.6	20.3
		12.5	26.0
MNU + SO	$0.6 \pm 0.2(4)$	0.9	$0.8 \pm 0.5(4)$
		0.3	
Citrate + PH	$0.2 \pm 0.1(4)$	0.3	$0.5 \pm 0.3(4)$
		0.5	
MNU, no operation	0.7	—	—
	0.4		
No MNU, no operation	—	$0.6 \pm 0.2(5)$	—

For scoring purposes, we used the characteristic histochemical staining for γ -GT, as described by Rutenburg *et al.*¹⁸, with which the selected foci stain dark red and the surrounding normal hepatocytes show no staining. γ -GT-positive foci were counted by low power microscopy and the numbers expressed as foci cm^{-2} of section area (mean $\pm$ s.e.m.). The number of animals is in parentheses. All the groups were selected by exposure to 0.02% 2-AAF plus CCl_4 or α -HCH, as detailed in the legend to Fig. 1.

The observation that cell replication seems to be necessary for the induction of putative initiated hepatocytes by MNU suggests that initiation of carcinogenesis may consist of at least two steps: (1) interaction of a carcinogen or its activated derivatives with target molecules, possibly DNA, to induce an appropriate biochemical lesion, and (2) the 'fixation' of the lesion or lesions by a round of cell replication. Such a concept would imply that one or more cellular events closely related to cell replication are an integral part of the initiation process with chemical carcinogens. The similarity between initiation of cancer development with chemicals and transformation of cell cultures with DNA viruses and X rays¹⁹⁻²¹ is noteworthy in that they both require at least one round of cell proliferation. The possibility that the initiating carcinogen is selecting a preformed altered cell rather than inducing a new one *de novo* is thought to be unlikely because of the following observations. (1) No foci of altered cells are selected without previous exposure to the carcinogen. (2) The foci observed are distributed randomly throughout the hepatic parenchyma, despite the striking zonal organisation of the liver. In addition, the demonstration that a carcinogen can permanently alter the progeny of a single cell in culture²² supports the hypothesis that the foci we observe represent a *de novo* induction phenomenon.

The possible role of altered DNA or other macromolecules and of the repair of such damage is now being analysed with this model.

This work was supported by the NCI of Canada, the US NCI and the MRC Canada ME-5994 and CONICIT, Venezuela.

EDMUNDO CAYAMA
HIROYUKI TSUDA
D. S. R. SARMA
EMMANUEL FARBER

University of Toronto,
Department of Pathology,
Medical Sciences Building,
Toronto, Ontario M5S 1A8, Canada

Received 8 May; accepted 24 July 1978.

- Craddock, V. M. in *Liver Cell Cancer* (eds Cameron, H. M., Linsell, D. S. & Warwick, G. P.) 153-201 (Elsevier, Amsterdam, 1976).
- Rajewsky, M. F. *Z. Krebsforsch.* **78**, 12-30 (1972).
- Warwick, G. P. *Fedn Proc.* **30**, 1760-1765 (1971).
- Foulds, L. *Neoplastic Development* **1**, Ch. 3 (Academic, London, 1969).
- Farber, E. *Cancer Res.* **33**, 2537-2550 (1973).
- Solt, D. & Farber, E. *Nature* **263**, 701-703 (1976).
- Solt, D., Medline, A. & Farber, E. *Am. J. Path.* **88**, 595-610 (1977).
- Cameron, R., Kellen, J., Kolin, A., Malkin, A. & Farber, E. *Cancer Res.* **38**, 823-829 (1978).
- Fiala, S., Fiala, A. E. & Dixon, B. J. *natn. Cancer Inst.* **48**, 1393-1401 (1972).
- Kalengayi, M. M. R., Ronchi, G. & Desmet, V. J. *natn. Cancer Inst.* **55**, 579-588 (1975).
- Swann, P. F. & Magee, P. N. *Biochem. J.* **110**, 39-47 (1968).
- Craddock, V. M. & Frei, J. V. *Br. J. Cancer* **30**, 503-511 (1974).
- Schulte-Hermann, R. *Cancer Res.* **37**, 166-171 (1977).
- Kaufman, D. G., Kaufman, W. K., Rice, J. M. & Wenk, M. L. *Proc. Am. Ass. Cancer Res.* **19**, 183 (1978).
- Rajalakshmi, S. & Sarma, D. S. R. *Chem.-Biol. Interactions* **11**, 245-252 (1975).
- Abanobi, S. E., Mulivor, R. A., Rajalakshmi, S. & Sarma, D. S. R. *Proc. Am. Ass. Cancer Res.* **17**, 103 (1976).
- Zahner, A. J., Rajalakshmi, S. & Sarma, D. S. R. *Proc. Am. Ass. Cancer Res.* **18**, 19 (1977).
- Rutenburg, A. M. *et al. J. Histochem. Cytochem.* **17**, 517-526 (1969).
- Borek, C. & Sachs, L. *Proc. natn. Acad. Sci. U.S.A.* **57**, 1522-1527 (1967).
- Borek, C. & Sachs, L. *Proc. natn. Acad. Sci. U.S.A.* **59**, 83-85 (1968).
- Todaro, G. J. & Green, H. *Proc. natn. Acad. Sci. U.S.A.* **55**, 302-308 (1966).
- Mondal, S. & Heiderberger, C. *Proc. natn. Acad. Sci. U.S.A.* **65**, 219-225 (1970).

C-type viral antigen expression in human placenta

THE evidence of C-type virus expression in man includes electron microscopic observation of virus-like particles in 30-50% of a small number of human placentas¹⁻⁴. Because of the limitations of ultrastructural surveys (sample number, specimen size, false negatives) we attempted to detect C-type virus expression in human placenta by indirect immunofluorescence using antisera raised against HEL-12 virus; this is a C-type retrovirus recovered from human embryonic lung cells after *in vitro* cultivation⁵. We demonstrate here C-type virus-related antigen

in 24 of 24 term placentas. The specificity of the immunofluorescence reaction for viral antigen was confirmed by serum absorption tests with purified virus. Preliminary isolation of the reactive antigen from placenta yielded a glycoprotein-rich fraction (β - μ) which blocked binding of anti-HEL-12 virus antiserum with placenta. In addition, antisera raised against β - μ reacted with HEL-12 virus-infected cells as well as with human placenta.

A survey of 24 normal full-term placentas was carried out with anti-HEL-12 virus antiserum. The HEL-12 virus antiserum was prepared by hyperimmunisation of rabbits with gradient-purified HEL-12 virus. The antiserum reacts with HEL-12 virus infected cells of several species but not with uninfected controls⁶. In addition, anti-HEL-12 virus antiserum has not been found to react with normal adult tissues^{6,7}. Comparable samples from the maternal surface of each placenta were quick-frozen in isopentane, usually within 1 h and no longer than 5 h after delivery, and then sectioned. Using a previously described technique⁶, all placentas exhibited the same distinctive pattern of fluorescence with anti-HEL-12 virus antiserum (Fig. 1a). For comparison, a light micrograph of a serial section of the same placenta is presented in Fig. 1b. Arrows indicate the areas in which immunofluorescence was detected. Fluorescence was predominantly associated with cells

Fig. 1 Immunofluorescent reaction in human placenta P28V. *a*, 4 μ m cryostat sections of human placenta (P28V) were washed in PBS and fixed in ether-ethanol (1:1) for 5 min followed by 95% ethanol for 20 min at 25 °C. Sections were then incubated with anti-HEL-12 virus antiserum (1:20) for 30 min in a moist environment at 25 °C, washed three times with PBS, stained with goat-anti-rabbit IgG-FITC conjugate (Behring) for an additional 30 min, and then washed again three times with PBS. Slides were examined with UV illumination using an Olympus BHA fluorescence microscope fitted with an FITC filter and a vertical ultraviolet illuminator. The reaction shown is typical of all of the placentas examined. *b*, A serial section of P28V was fixed in formalin and then stained with haematoxylin and eosin. The syncytiotrophoblast (s), cells of the villous stroma (v) and position of maternal (m) and fetal (f) circulation are marked. The arrows indicate the areas in which immunofluorescence was detected.

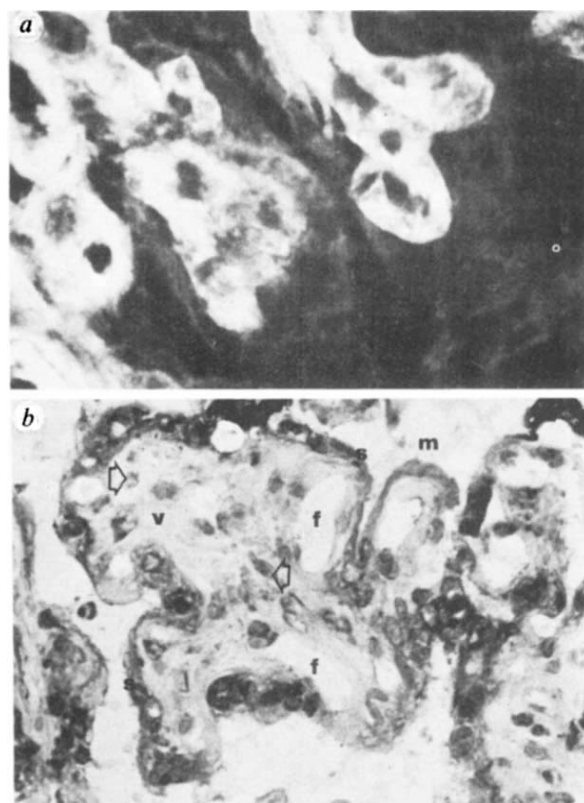


Table 1 Indirect immunofluorescence of human placenta with C-type viral antisera

Antisera (serum dilution)	Absorbent	Concentration (μ g protein)	Immunofluorescence	No. of placentas examined
Preimmune rabbit (1:20)	—	—	—	24
Rabbit anti-HEL-12 (1:20)	—	—	+	24
Rabbit anti-HEL-12 (1:20)	HEL-12 virus	25	—	5
Rabbit anti-HEL-12 (1:20)	SiSV	25	+	2
Rabbit anti-HEL-12 (1:20)	—	150	trace	2
Rabbit anti-HEL-12 (1:20)	BaEV	25	+	2
Rabbit anti-HEL-12 (1:20)	—	150	trace	2
Rabbit anti-HEL-12 (1:20)	β - μ (3 preparations)	40	—	4
Rabbit anti-HEL-12 (1:20)	RSV	150	+	4
Rabbit anti-HEL-12 (1:20)	Fetal calf serum	1,800	+	2
Rabbit anti-HEL-12 (1:20)	Bovine serum albumin	200	+	2
Rabbit anti-HEL-12 (1:20)	Human serum albumin	200	+	2
Rabbit anti-HEL-12 (1:20)	Haemoglobin	200	+	2
Rabbit anti-HEL-12 (1:20)	Human IgG	200	+	1
Goat-anti-SiSV (1:5)	—	—	+	2
Control goat (1:2)	—	—	—	2
Rabbit-anti-BaEV (1:2)	—	—	—	2

Indirect immunofluorescence of human placenta with C-type viral antisera. Antisera were used in indirect immunofluorescence tests⁶ with frozen sections of human placenta as described in the legend to Fig. 1, with or without previous absorption. Absorption was carried out incubating 50- μ l volumes of anti-HEL-12 virus antiserum (1:20) with 25–1,800 μ g of absorbent protein for 1 h at 25 °C and 18 h at 4 °C. Protein concentration was determined by the method of Lowry²². HEL-12 virus was purified as described⁵. Fetal calf serum (Gibco), human and bovine serum albumins (Pentex), haemoglobin (Sigma) and human IgG (Pentex) were diluted immediately before use. Secondary antisera were goat-anti-rabbit IgG-fluorescein isothiocyanate (FITC) and rabbit-anti-goat IgG-FITC antisera (Behring). Slides were evaluated under code by two independent observers. Each placenta specimen gave the reaction indicated. Specimens were negative (—), positive (+) or trace, a fluorescent reaction which was greatly reduced as compared with positive sample but which could be distinguished from negative samples.

of the villous stroma surrounding vessels carrying fetal blood. No significant fluorescence was found in syncytial trophoblast cells, which are in contact with maternal blood. The fluorescence pattern is consistent with what might be expected if the zone of antigenicity was delineated by the basement membrane between the syncytial trophoblast and the villous stroma. Neither the preimmune serum nor the fluorescein conjugate alone reacted with any placenta. Replicate sections of 10 placentas stained with twofold serial dilutions of antiviral antiserum showed a common endpoint of 1:40, suggesting that they contained a comparable amount of antigen.

Specificity of immunofluorescence was tested by serum absorption. Anti-HEL-12 virus antiserum was absorbed with gradient-purified simian sarcoma virus (SiSV), baboon endogenous virus (BaEV), Rous sarcoma virus (RSV) and HEL-12 virus. Table 1 shows that only HEL-12 virus blocked the reaction completely at the concentrations of virus used. SiSV and BaEV mediated partial absorption, as might be expected, for HEL-12 virus crossreacts with both although they do not crossreact with each other^{5,8,9}. RSV did not block the reaction. Bovine and human serum albumins, immunoglobulin and haemoglobin, at concentrations 10-fold in excess of that needed for HEL-12 virus to absorb, and a 70-fold excess of fetal calf serum, did not block the reaction. Antisera to SiSV and BaEV were tested for reactivity with placenta (Table 1). Both sera detect multiple components of HEL-12 virus as determined by immunoprecipitation and acrylamide gel analysis (J. T. Reynolds, unpublished observations). Anti-SiSV predominantly reacts with the viral envelope whereas anti-BaEV detects a wide range of virion components. Concentrated anti-BaEV antiserum did not react with placenta whereas anti-SiSV antiserum mediated a similar fluorescence pattern to that seen with HEL-12 virus antiserum. The data indicate that anti-HEL-12 virus antiserum reacts specifically with normal human placenta by detecting antigen which crossreacts with a primate C-type virion component. The inability of anti-BaEV antiserum to react with placenta compared with the effectiveness of concentrated BaEV in serum absorption tests with anti-HEL-12 virus antiserum most probably reflects a low antibody titre of the anti-BaEV antiserum for antigenic determinants shared by HEL-12 virus, BaEV and human placenta.

As C-type viral antigens are found at infected cell surfaces¹⁰, a procedure designed to elute cell surface components was used in an attempt to isolate the reactive placenta antigen¹¹. Reactive antigen was monitored during the isolation procedure by the ability of fractions to inhibit immunofluorescence in serum absorption tests. The chorionic membranes and umbilical cords were removed from placenta and the remaining tissue was homogenised and incubated with 3 M KCl for 18 h at 4 °C. The eluted material which was soluble in deionised water was chromatographed on G-100 Sephadex in phosphate-buffered saline (PBS), pH 7.3 (Fig. 2). Protein eluting between 70–80,000 MW, fraction 2, was the most active in serum absorption tests. This fraction was separated by chromatography with concanavalin A-coupled Sepharose (Con A) and a Con A binding fraction (β - μ) was eluted with α -methylmannoside. 40 μ g of this glycoprotein-rich fraction completely blocked the reaction of anti-HEL-12 virus antiserum with placenta (Table 1). In contrast, as 100 μ g of Sephadex fraction 2 was required for comparable inhibition, Con A chromatography selects for material which is increased in specific activity for the reactive

Table 2 Indirect immunofluorescence with anti- β - μ antiserum

Target cell	Infecting virus	Species	Immunofluorescence with antisera against HEL-12 Preimmune virus sera		
HEL-12 p. 21	HEL-12	human	+	+	—
A204	—	human	—	—	—
NMF	HEL-12	mink	+	+	—
NMF	—	mink	—	—	—
71AP1	SiSV	marmoset	+	+	—
MFS	—	marmoset	—	—	—
BILN	BaEV	baboon	+	+	—

Rabbit antisera were diluted 1:80 and used in indirect immunofluorescence tests with virus-infected or control cells grown *in vitro*²³. The secondary anti-serum, goat-anti-rabbit IgG-FITC (1:20) did not react with any cell by itself. Positive immunofluorescence with tissue culture cells was granular and apple green and confined to the cytoplasm. Slides were evaluated under code.

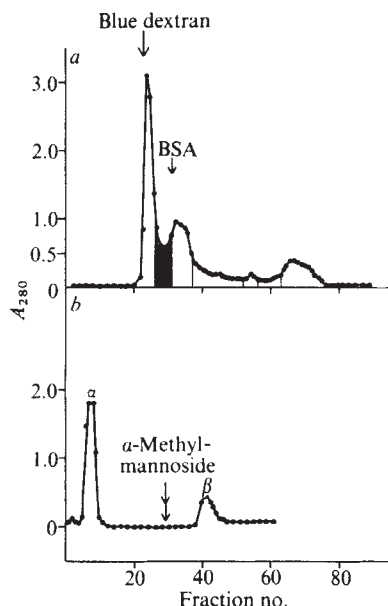


Fig. 2 Fractionation of KCl extract of human placenta. 200–300 g of placenta were obtained by removing the chorionic membrane from freshly delivered human placentas and washing the resultant material with PBS. Three 5-s homogenisations in a Waring blender were followed by treatment for 10 min with 0.85% NH_4Cl at 4 °C and the pellet obtained was treated with PBS containing 3 M KCl for 18 h at 4 °C. The resultant supernatant was extensively dialysed against 0.85% NaCl followed by deionised water. The water-soluble fraction was then concentrated by pressure dialysis and 20–50 mg in 1–1.5 ml was applied to a 150-ml Sephadex G-100 (Pharmacia) column (a). The column was run at a flow rate of 10 ml h^{-1} with PBS, 0.2% NaN_3 , pH 7.0, as the buffer. 2.0-ml fractions were collected and monitored at a wavelength of 280 nm in a Gilford 240 spectrophotometer. The column was calibrated with blue dextran, bovine serum albumin (BSA), and ribonuclease. Fractions were tested in serum absorption tests as described in Table 1. Based on such tests, the fraction shown shaded was selected and applied to a 10-ml column containing Con A-Sepharose (Pharmacia) (b). The column was then rinsed with four column volumes of normal saline, 0.05 M phosphate and 0.2% NaN_3 , pH 7.0, at a rate of 6.0 ml h^{-1} . This was followed by a 10% solution of α -methylmannoside (Sigma). 1.0-ml fractions were collected. The α -methylmannoside eluate was concentrated by pressure dialysis and used for serum absorption tests and immunisation.

placental antigen. β - μ preparations were isolated from three placentas and each preparation blocked the reaction of anti-HEL-12 virus antiserum with specimens from several different placentas.

To substantiate that β - μ contained antigen related to HEL-12 virus, an antiserum was raised in a rabbit against β - μ . Table 2 shows that this anti- β - μ antiserum reacted with virus-infected cells and human placenta but not with uninfected control cells. Preimmune serum did not react with any specimen.

Although we previously reported HEL-12 virus-related antigen and antibody in some tissue from patients with systemic lupus erythematosus^{6–9}, this is the first report of virus-related antigen in all specimens tested of a selected normal human tissue. The placenta may be a privileged site of C-type viral expression in man, as has been suggested for sub-human primates^{2,12}. Factors allowing placenta to be such a privileged site may include hormonal modulation of viral replication^{13,14} and the suspension of some immune mechanisms¹⁵.

The reaction of antiviral antiserum was predominantly associated with cells in contact with fetal blood whereas the rare particles reported in electron microscopic studies^{1–4} were found at the surface of cells bathed by maternal blood. This difference in the location of antigen and maturing particles is of interest in

view of the observation that mammalian cells often contain proviral DNA of non-crossreactive C-type viruses¹⁶. Four unrelated classes of C-type viruses have been described in primates^{17–19}. In placenta, the virions detected morphologically and the antigen detected by fluorescence may be related. The apparent absence of fluorescence in the syncytiotrophoblast may indicate that viral expression is restricted to small focal zones whereas viral expression in the cells of the villous stroma is widespread. Furthermore, virions in the villous stroma may not yet have been detected. Alternatively, the virions and β - μ antigen may not be antigenically related. In addition, there is evidence that C-type viral antigen may be transiently expressed during gestation^{20,21}.

We have consistently detected the presence of a component of human placenta which crossreacts antigenically with the C-type retrovirus HEL-12. We are able to isolate a fraction from placenta which inhibits this reaction of anti-HEL-12 virus antiserum with placenta and an antiserum prepared to this antigen reacts with placenta as well as with HEL-12 virus-infected cells. We have not determined the site of β - μ synthesis, nor do we know the chronology of its expression during gestation. Experiments are in progress to explore these questions as well as to characterise the complexity of β - μ and determine its presence in fetal tissue.

This work was supported by the Elsa U. Pardee Foundation, a Basil O'Connor Starter research grant from the NF, March of Dimes and CA 19264 from the NCI. We acknowledge the assistance of Drs P. Graff, N. G. Ordóñez, W. H. Kirsten and E. Kieff and Ms H. Dalton. RSV and anti-SiSV antiserum were obtained through the Office of Program Resources and Logistics, NIH.

M. H. SAWYER
N. E. NACHLAS, JR
S. PANEM

Department of Pathology,
and
Committee on Virology,
Pritzker School of Medicine,
University of Chicago,
Chicago, Illinois 60637

Received 20 March; accepted 25 July 1978.

- Kalter, S. S. *et al.* *J. natn. Cancer Inst.* **50**, 1081–1084 (1973).
- Dalton, A. J. *et al.* *J. natn. Cancer Inst.* **52**, 1379–1381 (1974).
- Imamura, M., Phillips, P. E. & Mellors, R. C. *Am. J. Path.* **83**, 383–389 (1976).
- Dirksen, E. R. & Levy, J. A. *J. natn. Cancer Inst.* **54**, 1187–1192 (1977).
- Panem, S. *et al.* *Science* **189**, 297–299 (1975).
- Panem, S. *et al.* *New Engl. J. Med.* **295**, 470–475 (1976).
- Ordóñez, N. G. *et al.* *Virchow Arch. path. Anat. Physiol.* **25**, 355–366 (1977).
- Prochownik, E. V. & Kirsten, W. H. *Nature* **260**, 64–67 (1976).
- Prochownik, E. V. & Kirsten, W. H. *Nature* **267**, 175–177 (1977).
- Panem, S. & Schauf, V. *J. Virol.* **13**, 1169–1175 (1974).
- Meyer, A. thesis, Univ. Chicago (1977).
- Kalter, S. S. *et al.* *J. natn. Cancer Inst.* **55**, 735–736 (1976).
- Schaller, J. P. *et al.* *Cancer Res.* **36**, 1980–1990 (1976).
- Wu, A. M. *et al.* *J. Virol.* **14**, 802–812 (1975).
- Bur, A. E. & Billingham, R. E. *Adv. Immun.* **14**, 1–84 (1971).
- Aaronson, S. A. & Stephenson, J. R. *Biochim. biophys. Acta* **458**, 323–354 (1976).
- Theilen, G. H. *et al.* *J. natn. Cancer Inst.* **47**, 881–885 (1971).
- Todaro, G. J. *et al.* *Proc. natn. Acad. Sci. U.S.A.* **75**, 1004–1008 (1978).
- Todaro, G. J. *et al.* *Cell* **13**, 775–782 (1978).
- Strand, M., August, J. T. & Jaenisch, R. *Virology* **76**, 886–890 (1977).
- Piko, L. *Cell* **12**, 697–707 (1977).
- Lowry, O. H. *et al.* *J. biol. Chem.* **193**, 265–275 (1975).
- Panem, S. *et al.* *J. gen. Virol.* **35**, 487–495 (1977).

Induction of extensive fimbrial branching in the adult rat brain

IT is now well known that some central nervous system axons can regenerate following axotomy^{1,2} or sprout in reaction to nearby denervation³. The cholinergic projection from the septal nuclei to the hippocampal formation has demonstrated the ability both to regenerate^{4,5} and sprout^{6,7}. However, the possibility that axons in the mature mammalian brain can be induced to grow in other circumstances has not been investigated. We have examined the ability of the axons of the

fimbria, the fibre bundle which interconnects the septal nuclei and the hippocampal formations, to grow in circumstances other than those used in classical regeneration and sprouting research. The fimbria is particularly suitable for the study of axonal growth in brain as its fibres run as an unbranched bundle, with one surface exposed to the lateral ventricle and thus free from surrounding neuropil. Consequently, it is readily accessible for experimental manipulation. We demonstrate here that fibres of the fimbria in the adult rat brain can be induced to grow in circumstances that neither sever them nor denervate regions to which they normally project.

We used adult albino rats (380–470 g, Sprague-Dawley from Simonsen Labs). All rats, under Nembutal anaesthesia, had their right fimbria exposed at the juncture of the septum, fimbria and anterior hippocampus. A flap of skull was removed for 2 mm on either side of the landmark bregma, and extending from the mid-sagittal suture to the edge of the dorsal surface of the skull. The dura was removed and the cortex, corpus callosum and caudate nucleus overlying and adjacent to the fimbria were aspirated. In some experiments the fimbrial fibres were left untreated, and in other experiments the fimbria was cut with a scalpel or treated for 10 min with a cotton swab containing either 1 μ g of NaCl or 1 μ g of colchicine. The wound was then rinsed with sterile saline and filled with Gelfoam. About two months after surgery animals were anaesthetised and perfused transcardially with a 10% formaldehyde-saline solution (a few animals were perfused with glutaraldehyde for ultrastructural observations). The brains were removed and post-fixed in the perfusion solution before staining and histological analysis. Additional animals were killed at 4 and 11 d after treatment to determine short-term effects of the experimental treatment.

We have previously found that after the fimbria has been cut or crushed the fibres do not regenerate. However, a less traumatic lesion, by freezing the fimbria, produced a most unusual outcome. In some animals, following freeze lesions of the fimbria, a bridge containing nerve fibres and glial and epithelial cells was seen to originate from the fimbria, cross the ventricular space and make contact with the lesioned caudate nucleus⁵ (Fig. 1*a, b*). These seemed to be regenerating fibres that for some reason could not pass through their normal trajectory and instead followed a path towards the caudate nucleus. However, during experiments on the trophic regulation of synaptogenesis in the mature rat brain we observed that the cotton swab application of colchicine to the fimbria often induced an outgrowth of fimbrial fibres. In some cases these fibres bridged the ventricular space between the fimbria and caudate nucleus. Initially, we suspected that these fibres had regenerated following colchicine-induced damage of the fimbria; however, this interpretation was unlikely in view of morphological evidence that the cotton swab-colchicine application caused very little, if any, fibre destruction (D.G. and C.W.C., in preparation). Accordingly, we proceeded to determine if the application of a cotton swab with salt crystals, or the mere surgical procedure of exposing the fimbria, could induce fimbrial fibres to grow towards the lesioned caudate nucleus.

An outcropping of fibres from the fimbria was seen 30–60 d after cotton swab-NaCl treatment of the fimbria in 13 of 15 cases, and in 5 of these cases there was also an outgrowth of fibres from the lateral septum. These fibres arced out from the fimbria (or septum) but usually curved back and followed a normal trajectory. In material stained for acetylcholinesterase (AChE) a single fibre could be observed to depart from its normal linear trajectory and curve towards the caudate among a tangle of other axons and glial cells. Many axolemmal bulges or varicosities were present in these axons (Fig. 3*a*). Such varicosities were rarely seen in the normal fimbria. In five of these 15 cases fibres were observed to travel in a dense glial matrix across the lateral ventricle where they contacted the partially damaged caudate (Fig. 2*a, b*). The marked enrichment of glial cells in the branch was not seen in the normal fimbria.



Fig. 1 *a*, Photomicrograph of fibre-stained, horizontal section of rat brain 60 d after freezing of fimbrial fibres. Paraffin sections were stained with the Holmes' silver method¹² to demonstrate nerve fibres. The branch is seen to extend from the fimbria (*f*) to the caudate nucleus (*c*). Contralaterally, the fimbria and caudate nucleus are intact. Scale bar, 500 μ m. *b*, Enlarged portion of fimbrial branch to caudate nucleus showing nerve fibres travelling in a mesh of glial cells. Scale bar, 50 μ m.

The fibres in the branch clearly adjoined the caudate neuropil, but it was less clear whether they actually entered it. Ependymal cells lined the rostral surface of the branch and were apparently continuous with the ependymal lining of the fimbria and caudate.

In the initial stages of this apparent growth process, a damming up of AChE-positive material was seen at 4 d post-treatment. By 11 d post-treatment axons could be observed which deviated from their normal pathway and displayed swollen regions along their course (Fig. 3*b*). These changes were specific to the areas of cotton swab application.

The growth of fimbrial fibres towards the caudate did not seem to be related to the destruction of the fibres, as (1) at short times after treatment, staining for AChE remained normal in the hippocampus (see Fig. 2*a*) (in contrast to the marked depletion of AChE staining seen after fimbrial lesions⁸); (2) there was little apparent lethal fibre damage in the fimbria as determined by light and electron microscopy; (3) at 4 d after treatment, many normal axons were observed to travel laterally towards the caudate instead of taking the usual longitudinal course along the hippocampal formation; (4) in four cases of complete section of the fimbria no fibre outcroppings were observed.

The conditions for the initiation of growth seemed to depend on two factors. First, some type of mechanical disturbance seemed necessary to elicit growth. In 9 animals whose fimbriae were exposed but left untreated there was only one case where fimbrial fibres grew out towards the caudate. This is in contrast to the cotton swab-NaCl treatment of the fimbria where in 5 of 15 cases fimbrial fibres reached the caudate nucleus. However, in the extreme instance of freezing the fimbrial fibres, axonal growth occurred frequently but involved very few fibres. The most prolific growth occurred with cotton swab applications. It may be that the mechanisms of growth induction are common to all these cases but that in the mere exposure of the fimbria the disturbance is too small, whereas in the freezing of fimbrial

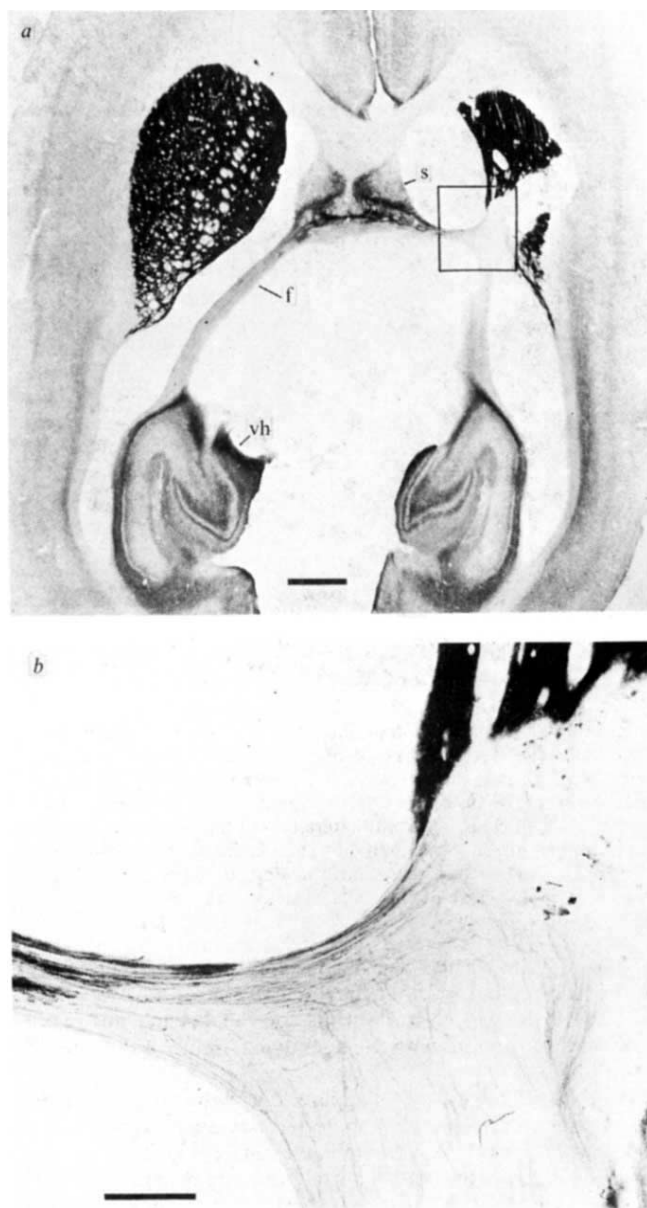


Fig. 2 *a*, Photomicrograph of AChE-stained, horizontal sections from rat brain 32 d after cotton swab-NaCl treatment of the fimbria. Section stained for AChE after the method of Matthews *et al.*¹³. The left side shows the normal course of fimbrial fibres travelling uninterrupted from the septum (s) to the ventral hippocampal formation (vh). The intact caudate nucleus is lateral to the fimbria. The right side shows the fimbria branching off towards the lesioned caudate nucleus. Although there is a disruption of fibres in the fimbria there is no depletion of AChE-staining in the ventral hippocampal formation. AChE-staining of the ventral hippocampal formation is almost completely eliminated by section of the fimbria⁸. In silver-stained material the branch is observed to contain many axons, of which only a small percentage stain positively for AChE. Calibration bar = 1 mm. *b*, Enlarged portion of boxed-in area of Fig. 2*a*, demonstrating AChE-positively staining fibres. Some of these fibres are shown to pass from the fimbria to contact the lesioned caudate in two separate regions, whereas other fibres continue posteriorly to the ventral hippocampal formation. Scale bar, 200 μ m.

fibres the extreme damage is a hindrance to growth. Second, the proximity of a lesioned caudate nucleus was apparently a condition for successful fibre outgrowth. In animals where most of the head of the caudate had been removed (the regions closest to the perturbed fibres of the fimbria) little, if any, growth of fimbrial fibres was observed. Other regions of close proximity to the fimbria, such as the thalamus, hypothalamus

and corpus callosum, were never targets for the outgrowing fibres.

We cannot specifically identify the nature of the perturbation that will initiate fibre growth. However, we feel that the cotton fibres from the swab, the mechanical disturbances of exposing the fimbria or the freezing of fimbrial fibres all act to irritate the fimbria and that this irritation induces nerve growth. The partially denervated cholinceptive cells of the caudate nucleus may produce substances which attract and stimulate the growth of cholinergic fibres of the fimbria that have been prompted to grow. The role of factors such as fibre perturbations and the attractive influence of denervated tissues in stimulating and sustaining axonal growth is very similar to those factors hypothesised by Cajal, Nageotte and Marinesco to be critical for growth of axons in the periphery and spinal cord of the mammalian nervous system⁹. Seiger and Olson have shown that locus coeruleus neurones that have matured in the eye can reinitiate extensive fibre growth in response to contact by a graft of denervated iris tissue, but not following contact with a host iris that still has innervation¹⁰. In these studies denervation was seen to be the important factor in stimulating fibre outgrowth from the *in oculo*, mature locus coeruleus neurones. Similarly, caudate denervation may act to stimulate fibre outgrowth from the fimbria.

Our finding that disturbed but non-lesioned nerves in brain can be induced to grow is very similar to results obtained by Cajal in the young cat after spinal cord damage⁹. In fact, the presence of axonal varicosities, as we observed in our AChE-stained fibres, was noted by Cajal to occur in regrowing spinal nerve fibres. Speidel reported collateral growth in normal peripheral nerves of the tadpole¹¹. These nerve fibres sometimes formed entirely new pathways, induced by the denervation of a nearby region. Speidel also noted that this type of axonal growth seemed to be characteristic of unmyelinated axons. In this respect, Svendgaard *et al.*⁴ have suggested that the unmyelinated fibres of the fimbria have an ability to regenerate that may be shared by all unmyelinated axons in the central nervous system. They report a regenerative capacity of these nerves, although disturbance-induced growth (and not axotomy-induced growth) of the cholinergic fimbrial fibres could also account for their observations. Instances of neuronal remodelling that occur following regeneration, sprouting and growth evoked by mechanical perturbation as described here, may all rely on similar growth-inducing factors.

Our findings suggest that the consequence of denervation can extend beyond the immediate vicinity of damage and apparently stimulate the growth of axonal branches over a

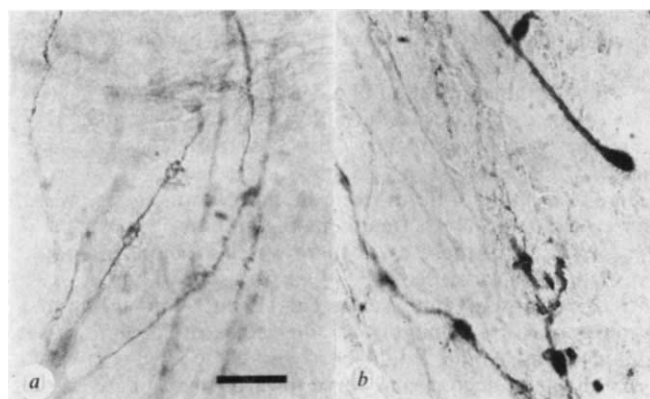


Fig. 3 *a*, Photomicrograph of AChE-positive axonal varicosities in fimbria branch 32 d after cotton swab-NaCl treatment of the fimbria. In cresyl-violet stained material these varicosities did not exhibit any staining. *b*, Photomicrograph of AChE-positive axonal varicosities 11 d after cotton swab treatment of the fimbria. Scale bar, 50 μ m.

distance a few millimetres from the denervated zone. This growth can apparently be initiated by several seemingly nonspecific mechanical disturbances of these nerves. Axonal growth from the fimbria may provide an opportunity to study the factors that effect initiation, stimulation and guidance of growing nerve fibres in the mature mammalian brain.

We thank Julene Mueller for secretarial assistance. This work was supported by NIH research grant MH 19691. D.G. was an NIMH predoctoral fellow (MH 0530801).

DAN GOLDOWITZ
CARL W. COTMAN

Department of Psychobiology,
University of California,
Irvine, California 92717

Received 27 April; accepted 26 June 1978.

- Guth, L. *Expl Neurol.* **48**, 3–15 (1975).
- Kerr, F. W. L. *Expl Neurol.* **48**, 16–31 (1975).
- Cotman, C. W. & Lynch, G. S. in *Neuronal Recognition* (ed. Barondes, S.) 69–108 (Plenum, New York, 1976).
- Svendgaard, N.-A., Björklund, A. & Stenevi, U. *Brain Res.* **102**, 1–22 (1976).
- Cotman, C. W., Hyatt, H., Kaups, P. & Lynch, G. *Neurosci. Abstr.* **1**, 499 (1975).
- Lynch, G., Matthews, D. A., Mosko, S., Parks, T. & Cotman, C. *Brain Res.* **42**, 311–318 (1972).
- Cotman, C. W., Matthews, D. A., Taylor, D. & Lynch, G. *Proc. natn. Acad. Sci. U.S.A.* **70**, 3473–3477 (1973).
- Lewis, P. R. & Shute, C. C. D. *Brain* **90**, 521–540 (1967).
- Ramon y Cajal, S. *Degeneration and Regeneration in the Nervous System* (ed. May, R. M.) I & II (Hafner, New York 1960).
- Seiger, A. & Olson, L. *Expl Brain Res.* **29**, 15–44 (1977).
- Speidel, C. C. *J. comp. Neurol.* **61**, 1–80 (1935).
- Carleton, H. M. & Drury, R. A. B. *Histological technique* 3rd edn, 229–231 (Oxford University Press, London 1957).
- Matthews, D. A., Nadler, J. V., Lynch, G. S. & Cotman, C. W. *Dev Biol.* **36**, 130–141 (1974).

Positive inotropism of vanadate in cat papillary muscle

MANY samples of 'Sigma grade' ATP from Sigma Chemical Co. reportedly contained an impurity which induces anomalous kinetics of (Na,K)-ATPase activity^{1–3}. This impurity has recently been identified as vanadate⁴, and shown to inhibit (Na,K)-ATPase from kidney⁴ and red blood cells⁵. These findings have received considerable interest because of a possible physiological role of vanadate as an endogenous regulator of (Na,K)-ATPase activity^{3–5}. We now report that vanadate is capable of producing positive inotropic effects in electrically driven papillary muscles isolated from cats.

Cats (0.5–1.6 kg) were anaesthetised with sodium pentobarbital (30 mg per kg i.p.) and papillary muscles (mean cross-sectional area 0.62 ± 0.05 mm², $n = 17$) were dissected from the right ventricles. The preparations were attached to a platinum stimulating electrode and mounted individually in glass tissue chambers for recording isometric contractions as described previously⁶. The bathing solution (5 ml) containing (mM) NaCl 136.9, KCl 5.4, CaCl₂ 1.8, MgCl₂ 1.05, NaH₂PO₄ 0.42, NaHCO₃ 11.9, glucose 5.5 was equilibrated with 95% O₂ + 5% CO₂ and maintained at 35 °C. The preparations were driven electrically at a frequency of 0.2 Hz (duration 5 ms, intensity about 10% above threshold). Drugs used were ammonium vanadate anhydrous (NH₄VO₃), sodium vanadate anhydrous (NaVO₃; both from Merck, Darmstadt) and (±)propranolol HCl (ICI). The compounds were freshly dissolved in bathing medium and injected into the tissue chamber in volumes of 300 µl or less. NH₄VO₃ and NaVO₃ did not contain any measurable calcium and did not change the pH (7.4) of the bathing solution.

The effects of NH₄VO₃ on force of contraction are illustrated by the representative experiment shown in Fig. 1. NH₄VO₃ (20–500 µM) was added cumulatively, and the time of exposure

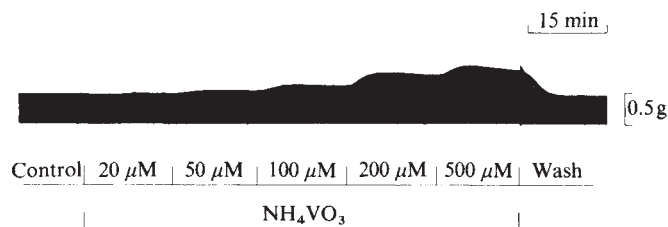


Fig. 1 Effect of NH₄VO₃ on isometric force of contraction of a cat isolated papillary muscle driven electrically at a frequency of 0.2 Hz. NH₄VO₃ was applied cumulatively at concentrations of 20–500 µM as indicated. The time of exposure to each concentration of NH₄VO₃ was about 15 min. After exposure to 500 µM NH₄VO₃, the preparation was washed three times with drug-free bathing solution. Temperature 35 °C. Extracellular Ca²⁺ concentration 1.8 mM. Preparation diameter 0.7 mm. Experiment 13 06 784. Similar results were obtained in each of nine other preparations.

to each concentration was about 15 min. NH₄VO₃ produced a concentration-dependent increase in force of contraction. The effect became clearly visible at 50 µM and was maximal at 500 µM. These concentrations correspond to those of Na₃VO₄ required to inhibit rubidium uptake in intact red cells by 50 to 80%⁵. The typical time course of action is also well shown. An increase in force of contraction was detectable after about 1 min, and a peak effect was attained after 6–8 min. The effect of NH₄VO₃ was readily reversible within less than 15 min on washing with drug-free bathing solution. Higher concentrations of NH₄VO₃ (1 mM) produced arrhythmias as a sign of toxicity.

The mechanism of the positive inotropic effect of NH₄VO₃ remains to be elucidated. Similar results as with NH₄VO₃ were obtained with NaVO₃ and, moreover, the effect of NH₄VO₃ was also observed in the presence of 1 µM of the β-adrenoceptor blocking agent propranolol (data not shown). This indicates that the positive inotropic action of NH₄VO₃ was neither due to an effect of the NH₄⁺ ion nor to a stimulation of β-adrenoceptors. Experiments are in progress to determine whether the positive inotropic effect of vanadate is accompanied by an inhibition of myocardial (Na,K)-ATPase, and, if so, whether both effects occur at similar concentrations of vanadate.

This work was supported by grant Scho 15/8 from the Deutsche Forschungsgemeinschaft.

INGELORE HACKBARTH
WILHELM SCHMITZ
HASSO SCHOLZ

Abteilung III (Biochemische Pharmakologie),
Institut für Pharmakologie und Toxikologie,
Medizinische Hochschule Hannover,
3000 Hannover 61, FRG

ERLAND ERDMANN
WOLFGANG KRAWIETZ
GUNTHER PHILIPP

Medizinische Klinik I,
Klinikum Großhadern,
Universität München,
8000 München 70, FRG

Received 4 July; accepted 24 July, 1978.

- Charney, A. N., Silva, P. & Epstein, F. H. *J. appl. Physiol.* **39**, 156–158 (1975).
- Beaugé, L. A. & Glynn, I. M. *Nature* **268**, 355–356 (1977).
- Beaugé, L. A. & Glynn, I. M. *Nature* **272**, 551–552 (1978).
- Cantley, Jr, L. C. *et al J. biol. Chem.* **252**, 7421–7423 (1977).
- Cantley, Jr, L. C., Resh, M. D. & Guidotti, G. *Nature* **272**, 552–554 (1978).
- Meinertz, T., Nawrath, H. & Scholz, H. *Naunyn-Schmiedeberg's Archs Pharmac.* **293**, 129–137 (1976).

Study of the charge-state model for electrophoretic variation using isoelectric focusing of esterase-5 from *Drosophila pseudoobscura*

VARIOUS neutral models have been proposed to account for the observed levels of enzyme polymorphism in natural populations¹⁻⁵. One such model, the charge-state or ladder-rung model², has been suggested to be appropriate for the available electrophoretic data. The model assumes that the only amino acid substitutions detectable by electrophoresis would involve unit charge changes—substitution of a charged for an uncharged residue—and because of charge redundancy alleles would fall into phenotypic classes of equal net charge^{3,4}. Thus an electrophoretic variant would be a heterogeneous molecular entity, rather than the product of a unique allele. It was proposed that these charge classes would be evenly spaced in electrophoretic mobility, because unit charge changes would be reflected as unit jumps in mobility, but Johnson⁶ has questioned this on the grounds that no two charged group substitutions should result in identical net charge changes. He argued that there would be no charge redundancy and that there would be a continuum of charged variants. It was also recognised that while many substitutions involving uncharged residues would probably go undetected, some could cause conformational changes which would be resolved by electrophoresis. This concern has been borne out by the electrophoretic data: although many enzymes fit the expectations for neutral alleles contingent on the charge-state model^{7,8}, others, including some esterases, do not. The mobilities of variants of these esterases are not evenly spaced,

and rare variants are often observed at intermediate as well as extreme electrophoretic mobilities. These characteristics have been attributed to the presence of conformation, as well as charge, variability. We have investigated the relative importance of charge and conformational change in producing electrophoretic variation by examining variants at the *Est-5* locus of *Drosophila pseudoobscura* by isoelectric focusing; this allows charge balance to be examined in the absence of conformational variation. The results indicate that while the observed charge changes fit a charge-state model, conformational differences exist which may account for a substantial amount of the additional variation revealed at this locus by conventional electrophoresis.

The use of variants of *Est-5* enabled us to examine charge balance in the absence of conformational difference by means of isoelectric focusing. The *Est-5* locus was first presented in Lewontin and Hubby's 1966 survey and characterised the extreme of allelic diversity for the set of loci that they examined⁹. Their list of six alleles was expanded to 12 in a later study of patterns of geographic variation¹⁰. *Est-5* has since been examined by additional electrophoretic methods which have more than doubled the former number of recognisable variant types¹¹⁻¹³.

We examined 22 electrophoretically distinguishable *Est-5* variants. Of these, 13 variants can be resolved using the original conditions of Lewontin and Hubby⁹ (hereafter termed standard conditions), while the other nine variants have been distinguished by other electrophoretic procedures^{11,13}. All the lines examined are either isogenic¹³ or electrophoretically homozygous¹¹ for the *Est-5* locus.

The relative positions of the 22 *Est-5* variants are shown in Fig. 1. These variants can be separated into only 10 groups on isoelectric focusing. Several variants which can be resolved by standard conditions of electrophoresis are indistinguishable from each other by isoelectric focusing, for example, 1.06 and 1.09 (Fig. 2). Of the nine variants which have been resolved by

Fig. 1 Relative positions of *Esr-5* variants after isoelectric focusing. The nomenclature used is the mobility of variants in standard electrophoretic conditions, 5% acrylamide gel pH 8.9 (ref. 9). Variants resolved by additional procedures are shown as follows: a, faster on 8% gels pH 8.9; b, slower on 8% gels pH 8.9; c, slower on 5% gels pH 7.1; d, faster on 8% gels pH 7.1; e, faster on 5% gels pH 10.7; f, slower on 5% gels pH 10.7 (all conditions, a-f, as ref. 13); g, faster on 14% starch pH 6.8 (ref. 11); M, enzyme is monomeric¹³; /, heterodimer enzyme

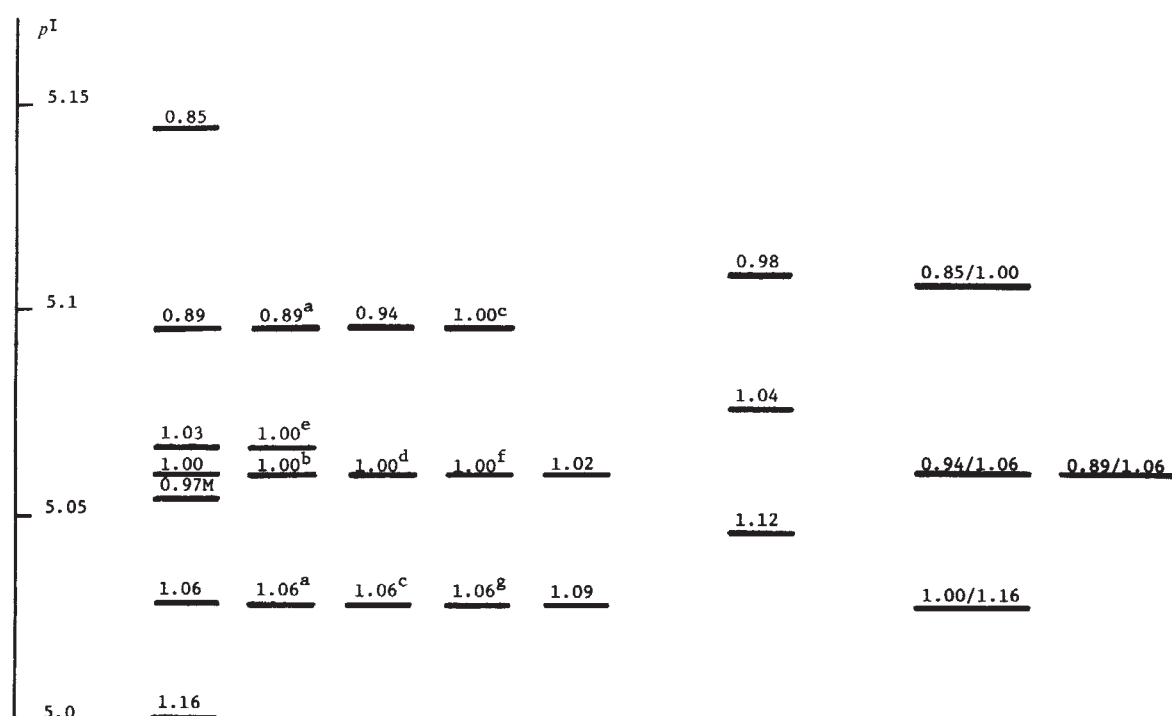


Table 1 Allocation of *Est-5* variants to charge classes

Charge class	Frequency*	<i>Est-5</i> variant
+2	0.012	0.85
+1	0.094	0.89, 0.89 ^a , 0.94, 1.00 ^c
0	0.457	0.97M, 1.00, 1.00 ^b , 1.00 ^d , 1.00 ^e , 1.00 ^f , 1.02, 1.03
-1	0.220	1.06, 1.06 ^a , 1.06 ^c , 1.06 ^g , 1.09
-2	0.027	1.16
Others		0.98, 1.04, 1.12

*Frequency values given are pooled frequencies of the individual variants as given by Prakash¹⁶.

non-standard electrophoretic conditions^{11,13}, only two, 1.00^c and 1.00^e, were resolved from the other variants in their standard classes by isoelectric focusing. One of these, 1.00^c, which is separable from other 1.00 variants by its slower mobility on 14% starch gel at pH 6.8 (ref. 11), occupied the same position as the 0.94 variant on isoelectric focusing.

Esterase-5 is a dimeric enzyme and the positions of some heterodimers are depicted in Fig. 1. Heterodimers are formed readily during electrophoresis of mixtures of homodimers. The 0.97 M variant behaves as a monomer in all conditions, and does not form stable heterodimers with the subunits of other variants.

The position of most variants after isoelectric focusing can be interpreted in terms of the charge-state model (Table 1). There are five main charge classes characterised by an essentially equal spacing. Three variants (0.98, 1.04, 1.12) exhibit very anomalous behaviour on isoelectric focusing and also do not fit the proposed charge classes; they are discussed below. The positions of the 1.16/1.00 and the 1.06/0.94 heterodimers show that these four variants are separated by equal charge changes. We propose that this is the result of substitutions which are approximately one unit charge jump apart. Small variations in charge are observed between the variants within single charge classes, but these variations in charge are small compared to the postulated unit charge jump. The apparent charge identity after isoelectric focusing of variants, which can be distinguished by standard conditions of electrophoresis (such as 1.06 and 1.09), suggests that differences due to apparent conformational changes are being detected by standard electrophoresis. Shifts in relative mobility could also represent small changes in net charge between variants being differentially expressed at the different pH conditions of electrophoresis.

The pooled frequencies of the variants within each major charge class fit a charge class configuration expected for neutral alleles with the central classes being most frequent (Table 1).

As pointed out above, the 1.00^c variant is not resolved from the 0.94 variant using isoelectric focusing. It is postulated that a substitution, resulting in an extra histidine residue, separates this variant from the other 1.00 variants. At the isoelectric point for *Est-5* (approximately pH 5.0), the 1.00^c variant would then jump charge classes; the size of this jump is consistent with our postulated unit charge change.

Considering their relative positions on isoelectric focusing, three variants (0.98, 1.04 and 1.12) display much faster mobilities in standard electrophoretic conditions than would be expected from their charges. These three variants lie approximately halfway between the postulated charge classes. They are evenly spaced and show the same charge jump increment as the major charge classes. These results suggest that they differ from each other by single charge changes, and may possess common features which separate them, as a group, from the variants which form the major charge classes.

Johnson¹⁴ has examined the charge-state model for *Est-5* indirectly by the use of gel-sieving. He claims that the combination of the free electrophoretic mobilities and retardation effects coincidentally result in apparent unit increments in mobility at pH 8.9. He concluded that the seven variants which he examined fit a charge class configuration; however, he did not study a

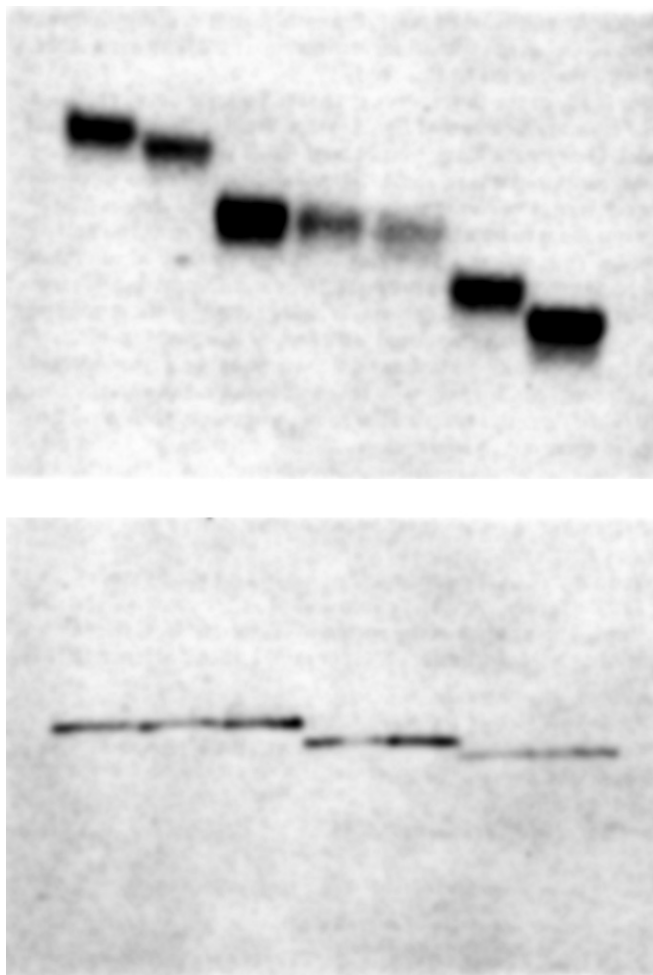


Fig. 2 Top: electrophoresis of (0.89, 0.94, 1.00^c, 1.00, 1.02, 1.06 and 1.09) *Est-5* variants in 5% acrylamide gel pH 8.9 (ref. 9). Bottom: isoelectric focusing of (0.89, 0.94, 1.00^c, 1.00, 1.02, 1.06 and 1.09) *Est-5* variants. Isoelectric focusing was based on the method of Vesterberg¹⁷, using 5% gels, 18 × 18 cm, polymerised by ammonium persulphate. Gels contained 1.6% w/v pH 4–6 and 0.4% w/v pH 3–10 ampholytes. Electrodes of Pt wire between Whatman 3MM paper were 16 cm apart; anodic and cathodic solutions were 0.2% v/v sulphuric acid and 0.4% v/v ethanolamine respectively. Samples, 10–15 μl, were loaded on 5 × 5 mm Whatman 3MM paper placed 2 cm from the cathode. Electrophoresis was at 4 °C for 4 h.; the voltage was gradually raised to 500 V over 1 h and maintained at this value. pH measurements were made at 20 °C on ampholytes eluted from transverse sections of gel. Staining for esterase activity was by the standard method¹³.

number of fairly common variants of intermediate mobility. Johnson claimed to find extensive genetic heterogeneity within allelic classes. He also found allelic heterogeneity in certain strains (some believed isogenic), yet no heterozygotes were found in these strains, which seems unlikely. More recently, Cobbs and Prakash¹⁵ examined the fit of *Est-5* to a charge state configuration using gel-sieving, and by chemically modifying the charge on several reference proteins. Using the average change in mobility of each reference protein per unit change in charge, they estimated an overall average mobility change per unit charge of 0.046 ± 0.029 . Chemical modification of charge on *Est-5* was not carried out. They assumed that the overall average was applicable to *Est-5*. As the magnitude of each charge change will vary with protein size and the proportion and balance of its charged residues, we consider it fortuitous that their estimate derived from diverse proteins was so close to our postulated unit charge change. The results of the studies by

Johnson¹⁴, and Cobbs and Prakash¹⁵, both using indirect methods of charge estimation, are incompatible with each other, and with the results of this study.

Our study of the differences in charge balance between *Est-5* variants shows clearly that *Est-5* can be fitted to the charge-state model. Despite the large numbers of *Est-5* variants segregating in natural populations, the total variability collapses into two offset charge class configurations, of five and three classes. There does not seem to be a continuum of charge change as claimed by Johnson⁶, but rather a clustering of charge variants around unit charge increments. Electrophoretic variation within charge classes and between the two major charge configurations may be the result of conformational differences exposed on electrophoresis. Our study indicates that conventional electrophoresis, in different conditions, will reveal considerably more genic variability than the simple charge heterogeneity shown by isoelectric focusing. Nevertheless, allelic variation at a complex locus such as *Est-5*, can be more clearly understood by the application of isoelectric focusing. Experiments are in progress to determine the structural basis for the differences between various *Est-5* variants.

We thank A. A. Felton and J. A. Coyne for the fly stocks, and R. D. J. Barker, A. A. Felton and R. C. Lewontin for discussions. This study was supported by a grant to R. C. Lewontin from the NIH Institute of General Medical Science (GM-21179).

JOHN A. M. RAMSHAW
WALTER F. EANES

Museum of Comparative Zoology,
Harvard University,
Cambridge, Massachusetts 02138

Received 22 March; accepted 19 July 1978.

- Kimura, M. & Crow J. F. *Genetics* **49**, 725-738 (1964).
- Ohta, T. & Kimura, M. *Genet. Res.* **22**, 201-204 (1973).
- King, J. L. *Genetics* **76**, 607-613 (1974).
- King, J. L. & Ohta, T. *Genetics* **79**, 681-691 (1975).
- Brown, A. H. D., Marshall, D. R. & Albecht, L. *Genet. Res.* **25**, 137-143 (1975).
- Johnson, G. B. *Genetics* **78**, 771-776 (1974).
- Bulmer, M. G. *Nature* **234**, 410-411 (1971).
- Richardson, R. H. & Smouse, P. E. *Biochem. Genet.* **14**, 447-466 (1976).
- Lewontin, R. C. & Hubby, J. L. *Genetics* **54**, 595-609 (1966).
- Prakash, S., Lewontin, R. C. & Hubby, J. L. *Genetics* **61**, 841-858 (1969).
- McDowell, R. E. & Prakash, S. *Proc. natn. Acad. Sci. U.S.A.* **73**, 4150-4153 (1976).
- Cobbs, G. *Genetics* **82**, 53-62 (1976).
- Coyne, J. A., Lewontin, R. C. & Felton, A. A. *Proc. natn. Acad. Sci. U.S.A.* (in the press).
- Johnson, G. B. *Genetics* **87**, 139-157 (1977).
- Cobbs, G. & Prakash, S. *Genetics* **87**, 717-742 (1977).
- Prakash, S. *Evolution* **31**, 14-23 (1977).
- Vesterberg, O. *Biochim. biophys. Acta* **257**, 11-19 (1972).

Table 1 Bile salt resistance of $\chi 1776$ in batch cultures

Inoculum size per 100 ml culture	Time after inoculation (h)	Proportion of resistant organisms	Culture pH
5×10^7 organisms	16	1 in 10^4 - 10^5	6.5
5×10^7 organisms	40	1 in 10 - 10^2	8.0
10^9 organisms	16	1 in 10 - 10^2	8.0

$\chi 1776$ was grown in L broth (1% Bacto-Tryptone, 0.5% yeast extract, 0.5% NaCl, 0.1% glucose, supplemented with 0.01% diaminopimelic acid, 0.004% L-threonine, 0.001% L-methionine, 0.0001% biotin and 0.004% thymidine). The initial pH of the culture medium was adjusted to 7.0 by addition of 1 M NaOH. The bile salt sensitivity of different cultures of $\chi 1776$ was monitored by plating out duplicate samples (0.2 ml) on L-agar and MacConkey agar supplemented with the specific growth requirements of the strain as given above. Essentially identical results were obtained with supplemented MacConkey agar plates containing 0.15% (Difco and Oxoid no. 3 formulations) and 0.5% (Oxoid no. 5 formulation) bile salts.

within experimental error, over a number of experiments but was dependent on the time of incubation at 37 °C, the initial inoculum size and the final pH of the culture supernatant (Table 1).

The variation in the culture pH indicated that the organism initially utilised the glucose present (0.1%) to produce acid. Depletion of glucose resulted in the utilisation of amino acids as carbon sources and in an increase in culture pH due to the secretion of ammonia. Differences in the numbers of bile salt-resistant organisms found here in primary cultures (low inoculum sizes and short incubation times) and those found by Curtiss (1 in 10^7 - 10^9) may be due to differences in the batches of medium used. Nevertheless, these results indicate that the bile salt sensitivity of $\chi 1776$ is dependent on culture conditions and that the surface properties of this organism might be subject to phenotypic change. The effect of growth rate on the bile salt sensitivity of $\chi 1776$ was therefore investigated in an attempt to gain further information. This was of additional interest as the mean generation time of *E. coli* in the mammalian intestine is considerably longer (12-24 h) than that observed using conventional laboratory techniques (0.5-1.5 h)^{3,4}.

A chemostat culture was set up so that $\chi 1776$ could be grown at a very low rate (Table 2). When steady-state conditions were reached (5 d) the culture was shown to be glucose limited (that is no glucose was present in the culture supernatant and addition of glucose to the chemostat produced a corresponding increase in cell density); the culture pH was uncontrolled and was 8.2. One in two of the organisms present were bile salt-resistant in these conditions. The dilution rate was then increased and at each steady state the number of bile salt-resistant organisms and the pH of the culture was measured. The proportion of bile salt-resistant organisms present in these different conditions remained approximately constant even though the culture pH, as expected, fell as the dilution rate increased (Table 2). This indicated that a bile salt-resistant variant had been selected.

In an additional experiment a chemostat culture of $\chi 1776$ was grown to steady-state conditions at a dilution rate of 0.025 h⁻¹, the culture pH being maintained at 6.0 by addition of 2 M HCl. The bile salt resistance of this culture remained at its starting level of 1 in 10^4 . The culture pH was then adjusted to 8.0 by addition of 2 M NaOH and the bile salt resistance monitored as a function of time. In less than one mean generation time, one in four of the organisms present had become bile salt resistant. This rate of change in the population in the chemostat indicates that the bile salt sensitivity of $\chi 1776$ is subject to phenotypic variation. After 10 d (during which the proportion of resistant organisms remained at one in four) the culture was returned to pH 6.0 by addition of 2 M HCl and a new steady state was established. Approximately one in ten organisms in the culture remained bile salt resistant, indicating that selection of a resistant variant had again occurred and that this variant was rela-

Bile salt sensitivity of *Escherichia coli* $\chi 1776$

Escherichia coli $\chi 1776$ ¹, a 'biologically contained' (EK2) host specifically constructed for genetic manipulation (ref. 1 and R. Curtiss, III, personal communication) contains many mutations which decrease its survival characteristics compared with wild-type *E. coli* and which also prevent its colonisation of the intestinal tract of mammals. The bile salt sensitivity of $\chi 1776$ is one of the parameters which preclude its survival in the intestine. This sensitivity is related to modifications of the cell envelope structure but, more specifically, to mutations at the *oms-1* or *rfb-2* loci. Curtiss (personal communication) has reported that the reversion frequency of $\chi 1776$ to bile salt resistance is approx. 10^{-7} to 10^{-9} . However, the cell surface structures of bacteria are subject to phenotypic variation and under control of a number of genes. In this paper we describe experiments which indicate that the bile salt sensitivity of $\chi 1776$ is phenotypically variable.

Batch cultures of $\chi 1776$ in L broth were monitored for bile salt resistance at 16 and 40 h after the initiation of growth. The number of bile salt-resistant organisms detected was constant,

tively stable. The growth of a single colony isolate of the bile salt-resistant variant in batch culture confirmed this stability. The variant grew well at pH 8.0, 7.1 and 6.3 in the presence of at least 0.25% bile salts and in the presence of nalidixic acid $50 \mu\text{g ml}^{-1}$. Furthermore, growth at pH 8.0 in the absence of bile salts produced a culture in which less than one in 10^4 organisms were bile salt sensitive. This increased to approximately one in 2×10^2 when the variant was grown at pH 6.3.

Throughout these experiments, lasting a number of mean generation times, no variation in the absolute requirements for diaminopimelic acid and thymidine were observed. Indeed, the cultures remained phenotypically $\chi 1776$ except for the variations detected in bile salt sensitivity. There had therefore been no contamination with other *E. coli* strains.

Table 2 Bile salt resistance of $\chi 1776$ in chemostat cultures

Dilution rate	Mean generation time (h)	pH of culture	Proportion of resistant organisms
Initial batch culture	1 (in log growth)	7.0→6.5	1 in 10^4
0.025	26	8.5	1 in 2
0.05	13	8.2	1 in 5
0.1	7	7.7	1 in 5
0.3	2	7.5	1 in 10
0.6	1.1	6.4	1 in 10
Wash-out point			

A Porton-type chemostat (500 ml L broth containing 0.1% glucose and growth supplements) was inoculated with $\chi 1776$ and this was grown at 37°C in batch culture conditions; after 16 h the chemostat was switched to the continuous culture mode with the dilution rate set at 0.025 h^{-1} (mean generation time 26 h). The pH was monitored but not controlled. The dilution rate was then increased and the number of bile salt-resistant organisms was determined at each new steady state. The culture density remained at approximately 5×10^9 – 10^{10} organisms ml^{-1} until wash-out.

These experiments do not necessarily preclude the use of $\chi 1776$ as a host in genetic manipulation. They do, however, show that, if care is not taken, selection for variants of this organism may occur. These variants, with slightly better growth characteristics, can quickly outgrow 'authentic' $\chi 1776$ produced by serial passage in batch cultures. In addition, bile salt-resistant variants of $\chi 1776$ have slightly better survival characteristics in animal tests and are better recipients for some conjugative plasmids (R. Curtiss, III, personal communication). These results indicate that any form of biological containment that is dependent on the surface properties of *E. coli* may be suspect.

Finally, and more importantly, these experiments demonstrate the need to monitor the safety features of any 'biologically contained' organism before, and during any genetic manipulation experiment.

This work was financed, in part, by a research grant from the Health and Safety Executive.

J. R. HUNTER
D. C. ELLWOOD
A. ATKINSON
P. J. GREENAWAY

Microbiological Research Establishment,
Porton Down,
Salisbury, Wiltshire, UK

Received 23 May; accepted 28 July 1978.

1. Curtiss, R., III *et al.* in *Recombinant Molecules: Impact on Science and Society* (eds Beers, R. F., Jr & Bassett, E. G.) 45–56 (Raven, New York, 1977).
2. Ellwood, D. C. & Tempest, D. W. *Adv. microb. physiol.* **7**, 83–117 (1972).
3. El-Shazly, K. & Hungate, R. E. *Appl. Microbiol.* **13**, 62–69 (1968).
4. Gibbons, R. J. & Kapsemalis, B. *J. Bact.* **93**, 510–512 (1967).

Expression of λ red genes restores *recB*-dependent protein X induction in *E. coli*

DNA DAMAGE and/or inhibition of DNA synthesis elicit a number of responses in *Escherichia coli* known as SOS functions¹, including the production of large quantities of a protein provisionally called protein X (ref. 2). Induction of protein X, like the other SOS functions, requires functional *recA* and *lexA* genes^{3,4}. Gudas and Pardee⁴ proposed a model to suggest roles for *recA* and *lexA* in the regulation of protein X production. Several research groups have identified protein X as the *recA* gene product^{5–8} which prompted them to formulate updated models, to account for the apparent autoregulation of the *recA* gene^{6–8}. These models propose that the *lexA* gene codes for a repressor which inhibits the transcription of the *recA* gene. Following treatment, such as exposure to ultraviolet irradiation or nalidixic acid, a small inducer molecule is produced that causes the *recA* gene product, present in low levels in untreated cells, to inactivate the *lexA* repressor. This allows transcription of the *recA* gene that results in the production of large amounts of the *recA* gene product, protein X. Gudas and Pardee⁴ suggested that the inducer molecule may be a DNA degradation product resulting from the action of the *recBC* gene products, exonuclease V. In support of this idea they showed that protein X is not induced by nalidixic acid in *recB*[–] mutants deficient in exonuclease V. On the other hand, Little and Hanawalt⁹ showed that intracellular degradation of unmodified phage DNA by restriction is not sufficient to induce protein X, and they have challenged the role of DNA degradation in the induction process. In this report we show that phage λ genes controlling λ exonuclease can restore the missing function for the induction of protein X in *recB*[–] mutants. This is further evidence that DNA degradation is required to derepress the *recA* gene.

As the exonuclease V protein has an ATPase activity in addition to both endo- and exonuclease activities¹⁰, the missing function for the induction of protein X in nalidixic acid-treated *recB*[–] mutants could be some function other than that of degrading DNA. We have therefore tried to determine whether λ exonuclease, an exclusively exonucleolytic enzyme¹¹, could circumvent the *recB* mutation in the induction of protein X. To do this we have determined whether protein X was induced in *recB*[–] lysogens that were transiently induced to produce λ exonuclease before treatment with nalidixic acid. In similar experiments others have shown that transient prophage induction enhanced *recB*-dependent cell survival after ultraviolet¹² or X-ray irradiation¹³.

Figure 1 shows the effect of transient heat induction of λ cI857 prophage on protein X induction in nalidixic acid-treated *recB*[–] lysogens. Lanes 8–12 show that the amount of protein X increased with increasing times of incubation of 42°C . In control experiments shown in lanes 3 and 4, neither a 5-min incubation at 42°C nor nalidixic acid treatment alone was sufficient to induce protein X. To investigate the role of λ exonuclease in this induction, the above experiments were repeated using *recB*[–] lysogens carrying prophage with mutations in genes controlling λ exonuclease activity. The combined treatment of heat induction and exposure to nalidixic acid was not sufficient to induce protein X when the prophage carried the mutations *redX314*, (lane 6) or *redB113* (lane 7). *RedX314* and *redB113* are point mutations in the genes which code for λ exonuclease and β protein, respectively^{14,15}. λ exonuclease acts primarily at 5'-phosphoryl termini on double-stranded DNA molecules, and the only acid-soluble products are deoxyribonucleoside 5'-monophosphates¹¹. The *redX314* mutation decreases the activity of λ exonuclease to less than 2% of normal activity¹⁴. Although the *redB113* mutation does not affect the λ exonuclease enzyme, it does alter the antigenic properties of the

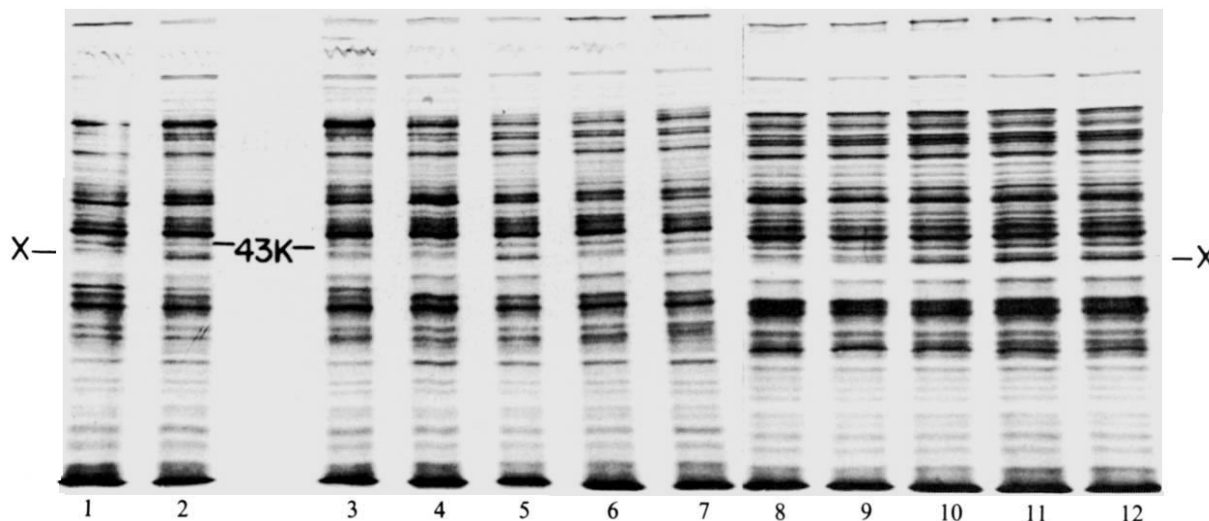


Fig. 1 Strain AB2470 *recB21* (ref. 21) was lysogenised with λ C1857, λ C1857 *redX314* or λ C1857 *redB113*. The λ C1857 mutant produces a temperature-sensitive repressor. The *redX314* and *redB113* mutations are described in the text. Lysogens were grown at 30 °C to an $A_{590} = 0.2$ in glucose-M9 media containing required amino acids and thiamine; 1-ml samples were incubated at 42 °C for various times before addition of nalidixic acid, 50 $\mu\text{g ml}^{-1}$ (Sigma). At 30 min after addition of the drug, ^{14}C -leucine (Schwarz-Mann) was added (to 5 $\mu\text{Ci ml}^{-1}$) and incubation was continued for 60 min at 30 °C. The cells were collected by centrifugation and prepared for SDS-polyacrylamide gel electrophoresis as described by Little and Hanawalt⁹. This figure is a composite picture of autoradiographs of two slab gels. Equal amounts of radioactivity were added to each slot in each gel. All lanes except 6 and 7 show proteins synthesised in AB2470 carrying λ C1857 prophage, unless otherwise noted, that were treated as described below. Untreated (lane 1); irradiated with 30 J m^{-2} UV (lane 2); incubated at 42 °C for 5 min (lane 3); exposed to nalidixic acid, 50 $\mu\text{g ml}^{-1}$ (lane 4); incubated at 42 °C for 5 min followed by exposure to nalidixic acid (lane 5); AB2470 (λ C1857 *redX314*) treated as in lane 5 (lane 6); AB2470 (λ C1857 *redB113*) treated as in lane 5 (lane 7); incubated at 42 °C for 1 min (lane 8), 2 min (lane 9), 3 min (lane 10), 4 min (lane 11) and 5 min (lane 12) followed by exposure to nalidixic acid.

β protein¹⁵. The only known activity of the β protein is to enhance the affinity of λ exonuclease for DNA *in vitro*¹⁶. This evidence indicates that the activity of λ exonuclease is responsible for the induction of protein X in nalidixic acid-treated *recB*⁻ lysogens. We interpret this to mean that DNA degradation is necessary for nalidixic acid-mediated induction of protein X.

As SOS functions, such as cell filamentation, error-prone DNA repair, and λ prophage induction, are dependent on the *recA*⁺ *lexA*⁺ genotype, the derepression of the *recA* gene seems to be necessary for their expression. For this reason, SOS functions are elicited by treatments that result in DNA degradation. How does the *recA* protein activate SOS functions? Roberts and Roberts¹⁷ have shown that λ prophage induction is due to a *recA*-dependent proteolytic cleavage of the λ repressor. They have recently found that the *recA* gene product, protein X, cleaves the λ repressor *in vitro* (J. W. Roberts, personal communication). Witkin¹⁸ proposed that all of the SOS functions are derepressed by a common mechanism, is by proteolytic cleavage of the repressor(s) controlling the SOS operon(s). DNA degradation, an initial consequence of DNA damage or of a stalled replication fork, may provide a primary danger signal which is amplified through the action of the *recA* protein resulting in the expression of a set of functions designed to enhance cell survival.

Although the above evidence clearly indicates that *recB*-dependent DNA degradation is required for nalidixic acid induction of protein X, this dependence may not apply to induction by other agents. For example, lane 2 of Fig. 1, indicates that ultraviolet induction of protein X is independent of the *recB* mutation, confirming the results of Little and Hanawalt⁹. Perhaps excision of ultraviolet photoproducts or post-replication repair results in sufficient DNA degradation, not requiring the *recBC* exonuclease, to trigger the derepression of the *recA* gene. In support of this suggestion, Braun¹⁹ has shown that ultraviolet induction of λ prophage does not occur if both excision and replication are eliminated genetically. Thus, we propose that more than one pathway of DNA degradation may lead to the induction of SOS functions.

What is the identity of the inducer? An attractive idea is that a modified DNA degradation product serves as a specific signal to communicate to the cell that its genetic material is in jeopardy, and that the SOS functions are needed. This would explain the results of Little and Hanawalt⁹ that degradation of unmodified phage DNA is not sufficient to cause induction. As modification is a post-replication process²⁰, a modified DNA degradation product could unequivocally be distinguished from DNA precursors normally present in the cell. Moreover, such a role for modification in the induction of SOS functions would emphasise the importance of pathways for the preservation of genetic material and would provide considerable insight into the cellular function of DNA modification.

This work was supported by grants from the American Cancer Society (ACS 77-073) and from the NIH (GM-19167).

ROBERT M. CROWL
SIRAJ UDDIN AHMED
RICHARD P. BOYCE

Department of Biochemistry and Molecular Biology,
University of Florida,
Gainesville, Florida 32610

Received 12 May; accepted 17 July 1978.

- Radman, M. in *Molecular Mechanism for Repair of DNA* (eds Hanawalt, P. & Setlow, R. B.) 355-367 (Plenum, New York, 1975).
- Gudas, L. J. & Pardee, A. B. *J. molec. Biol.* **101**, 459-477 (1976).
- Defais, M., Fauquet, P., Radman, M. & Errera, M. *Virology* **43**, 495-503 (1971).
- Gudas, L. J. & Pardee, A. B. *Proc. natn. Acad. Sci. U.S.A.* **72**, 2330-2334 (1975).
- Little, J. W. & Kleid, D. G. *J. biol. Chem.* **252**, 6251-6252 (1977).
- McEntee, K. *Proc. natn. Acad. Sci. U.S.A.* **74**, 5275-5279 (1977).
- Gudas, L. J. & Mount, D. W. *Proc. natn. Acad. Sci. U.S.A.* **74**, 5280-5284 (1977).
- Emmerson, P. T. & West, S. C. *Molec. gen. Genet.* **155**, 77-85 (1977).
- Little, J. W. & Hanawalt, P. C. *Molec. gen. Genet.* **150**, 237-248 (1977).
- Goldmark, P. J. & Linn, S. *J. biol. Chem.* **247**, 1849-1860 (1972).
- Little, J. W. *J. biol. Chem.* **242**, 679-686 (1967).
- Braun, A. & Gluck, D. *J. Bact.* **131**, 208-213 (1977).
- Trgovčević, Z. & Rupp, W. D. *Proc. natn. Acad. Sci. U.S.A.* **71**, 503-506 (1974).
- Radding, C. M. *J. molec. Biol.* **52**, 491-499 (1970).
- Shulman, M. J., Hallick, L. M., Echols, H. & Signer, E. R. *J. molec. Biol.* **52**, 501-520 (1970).
- Radding, C. M. & Carter, D. M. *J. biol. Chem.* **246**, 2513-2518 (1971).
- Roberts, J. W. & Roberts, C. W. *Proc. natn. Acad. Sci. U.S.A.* **72**, 147-151 (1975).
- Witkin, E. M. *Bact. Rev.* **40**, 869-907 (1976).
- Braun, A. *Nature* **261**, 164-166 (1976).
- Arber, W. *Progr. Nucleic Acids Res. molec. Biol.* **14**, 7-37 (1974).
- Emmerson, P. T. *Genetics* **60**, 19-30 (1968).

Haem exposure as the determinate of oxidation-reduction potential of haem proteins

KASSNER¹ has proposed that the apolarity of the haem environment primarily determines the oxidation-reduction potential, E'_0 pH 7.0, of a haem protein. Since high resolution crystallographic models are now available for six haem proteins with E'_0 values of between 20 and 320 mV, I have tested the Kassner proposal by comparing the kinds of atoms contacting the haem in each of these proteins. The results demonstrate that the apolarity of the total haem environment for these proteins is virtually identical but that the E'_0 value is inversely dependent on the exposure of the haem to an aqueous solvent.

Surface area calculations were made using a DEC PDP 11/45 computer and protein coordinates supplied by the Brookhaven Protein Data Bank. An unpublished computer program developed by Arnone and Briley (similar to that of Shrake and Rupley²) to both generate and display protein surfaces accessible to water solvent molecules was used. Each atom in a protein is represented by a sphere centred about its atomic coordinates with radius the sum of its Van der Waals radius and 1.4 Å, the radius of a water molecule. A three-dimensional grid of equally spaced points is constructed large enough to envelop the entire hydrated protein molecule. Grid points which lie within the space defined by spheres centred at each atom are termed interior points. The six nearest neighbours of each interior point are then examined. Those interior points which are not bounded on all sides by other interior points are identified as surface points. Calibration experiments have shown that the accessible surface of a sphere representing a typical hydrated protein atom is sampled by about 300 points using a 0.5 Å grid and results in a random calculation error of less than 5%. All surface points which lie on the surface of only one atom are assigned to that atom, while surface points shared by more than one atom are uniquely assigned to the atom having the nearest centre. The total number of surface points associated with each atom is converted to Å² of surface area using the factor 3.41 surface points Å⁻² obtained from a calibration curve. Carbon and sulphur atoms are considered as apolar atoms while all other haem protein atoms are considered as polar atoms.

All surface area calculations reported here were made using a 0.5 Å grid. The surface area of each haem and each apoprotein

was obtained by elimination of all coordinates associated with the apoprotein and the haem atoms, respectively, prior to calculation. The surface area of each haem crevice was calculated by insertion of the haem surface into the protein surface, removal of all points common to this composite surface and to the haem protein and assignment of the points remaining to individual atoms as above.

The total accessible surface area of the crystallographic model for each of the six haem proteins considered is listed in Table 1, entry 1. These surface areas are related to molecular weight by the expression, surface area = $11.0 M^{2/3}$, in excellent agreement with the equation developed by Teller³ using the surface areas calculated for proteins by the procedure of Lee and Richards.⁴ Approximately half ($53 \pm 3\%$) of the accessible surface area of these haem proteins is contributed by apolar atoms, entry 2, a percentage within the range calculated for five proteins by Lee and Richards⁵ and by Shrake and Rupley². These favourable comparisons indicate that our procedures for calculations of surface area and assignment of area to individual atoms are consistent with results obtained by others.

Removal of the coordinates for the apoprotein atoms from each of the haem proteins and subsequent surface area calculation, entry 4, indicates that a haem constructed from protoporphyrin IX has an accessible surface area of 805 ± 15 Å². As shown in Table 1, entry 3, between 77 and 96% of haem surface is buried in the haem proteins considered. Inspection of the haem atoms which are accessible in the haem proteins indicates (1) that neither the iron atom nor any of the pyrrole or methine bridge atoms are accessible, (2) that the atoms comprising some of the methyl, vinyl and propionate side chains are accessible, and, (3) that the haem is orientated in these haem proteins in one of two distinctive modes exposing different side chains as illustrated in Fig. 1. Haems which are not covalently attached to the polypeptide chains as in haemoglobin, myoglobin and cytochrome *b*₅, are orientated such that the side chain pyrrole rings 3 and 4 are accessible to the solvent, presumably to place the propionate carboxylates in a more polar environment. By contrast, haems which are covalently attached to polypeptides by addition of cysteinyl sulphhydryl groups across vinyl side chains, as in the cytochromes *c*, appear to be rotated by about 90° exposing the vinyl group of pyrrole 2 and the methyl groups of pyrroles 2 and 3. This orientation which buries the two propionates, would be expected to require complementary pairing of the buried carboxylate oxygen atoms by polar atoms in the haem crevice.

Table 1 Accessible surface area calculations of haem proteins

Entry	Parameter	Cyto <i>c</i> ₂	Cyto <i>c</i>	Cyto <i>c</i> ₅₅₀	Hb _α	Hb _β	Myo	Cyto <i>b</i> ₅
Haem protein								
1	Total accessible surface	5,473	5,307	6,613	6,770	7,152	7,120	4,858
2	Accessible apolar surface	3,019	2,860	3,414	3,849	3,760	4,000	2,352
3	Accessible haem surface	49	32	38	116	163	144	176
Haem								
4	Total accessible surface	796	819	832	800	805	799	783
Haem crevice of apoprotein								
5	Total accessible surface	740	782	784	677	632	645	593
6	Accessible apolar surface	504	550	623	534	516	509	474
7	Accessible aromatic apolar surface	200	140	158	150	146	137	118
Tabulations								
8	% Haem exposed in haem protein	6	4	5	14	20	18	23
9	% Haem crevice apolar	68	70	79	79	82	79	80
10	% Haem crevice aromatic	40	25	25	28	28	27	25
11	% Haem environment apolar	63	67	75	66	64	64	61
Reduction potential								
12	E'_0 , pH 7 (mV)	320 ⁶	260 ⁷	250 ⁸	113 ⁹	53 ⁹	47 ^{9,10}	20 ¹¹

The protein column headings reading from left to right refer to cytochrome *c*₂ from *Rhodospirillum rubrum*, ferricytochrome *c* from tuna heart, cytochrome *c*₅₅₀ from *Paracoccus denitrificans*, the α₁ subunit of human deoxyhaemoglobin, the β₂ subunit of that protein, sperm whale metmyoglobin, and cytochrome *b*₅ from bovine liver. Surface areas are expressed in Å².

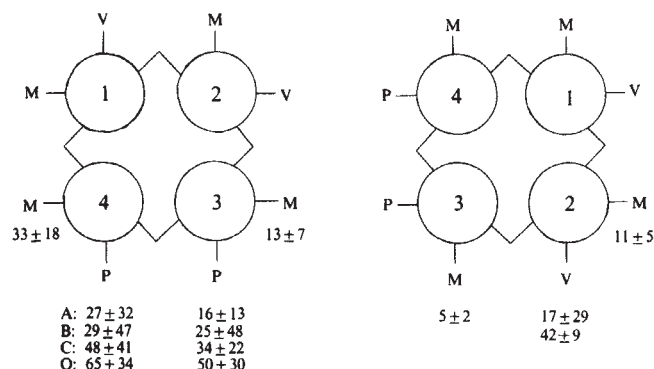
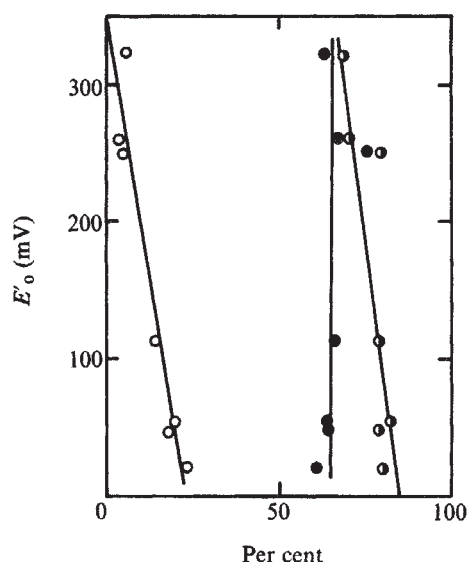


Fig. 1 Haem orientations in haem proteins. The haem present in haemoglobin, myoglobin and cytochrome b_5 is represented on the left and the haem in cytochrome c on the right. The pyrrole rings represented by circles numbered 1–4 and the pyrrole side chains represented by M, V and P for methyl, vinyl and propionate groups, respectively. Each haem is orientated such that its exposed side chains are at the bottom. The mean percentile exposure and standard deviation for the exposure of haem atoms in the haem proteins relative to their exposure in haem itself are listed below the atoms concerned. A, B, C and O refer to the exposure of the carbon atoms, A–C, and the oxygen atoms, O, of methyl, vinyl and propionate groups reading outwards from the pyrroles considered.

As shown in Table 1, at least 98% of the total accessible surface area for each haem, entry 4, can be accounted for by the sum of its accessible surface area in the haem protein, entry 3, and the accessible surface area of its crevice, entry 5. More than half, $61 \pm 15\%$, of the crevice polar surface is contributed by polypeptide backbone oxygen and nitrogen atoms. By contrast, only $8 \pm 3\%$ of the crevice apolar surface is contributed by polypeptide backbone carbon atoms. No good correlation exists between the E'_0 value and the percentage contribution of the atoms of any amino acid or groups of amino acids, for example aromatic apolar atoms, entry 10, to either the polar or apolar crevice surface. The apolar side chain atoms of Phe, Ile, Leu and Val residues contribute at least 50% of the crevice apolar surface while those of Asp, Glu, Asn, Gln and Arg contribute less than 8% in keeping with the occurrence of these residues in protein interiors⁵.

Fig. 2 Dependence of reduction potential, E'_0 , on the per cent exposure of the haem (○), on the per cent apolarity of the haem crevice (●) and on the per cent apolarity of the haem environment (●). All values were obtained from Table 1. Each line represents a least squares analysis of the values shown.



The dependence of the reduction potential on the haem exposure, on the apolarity of the haem crevice and on the apolarity of the haem environment (Table 1, entries 8, 9, and 11) are compared in Fig. 2. The burial of the haem is accompanied by a nearly parallel decrease in the apolarity of the haem crevice. This parallelsim seems reasonable since burial of one or both haem propionates and/or provision of the second extrinsic haem ligand by the protein would require complementary polar atoms in the haem crevice as noted above. The net effect of these parallel changes is to maintain the apolarity of the haem environment constant while the reduction potential varies by 300 mV as shown in Fig. 2. This observation is clearly contrary to the Kassner proposal.

As shown in Fig. 2, the reduction potential is inversely related to the fraction of haem surface exposed according to the expression $E'_0 = -15 (\% \text{ exposure}) + 345 \text{ mV}$. This empirical relationship may reflect the accessibility of the haem to environmental O_2 or the proximity of the extrinsic ligands to the haem iron. As more of the protein envelops the haem, the ferrohaem would be more protected from oxidants and/or the haem iron-extrinsic ligand bond distance would be diminished. Both of these changes should increase the reduction potential of a haem protein¹².

The equation relating E'_0 and haem exposure merits additional comment. This equation predicts that a totally exposed haem would have an E'_0 value of $-1,155 \text{ mV}$ and that a totally buried haem a value of $+345 \text{ mV}$. This range encompasses the E'_0 values reported^{13,14} for the majority of haem proteins except for cytochrome oxidase, cytochrome f and a few cytochromes c_{553} and c_{552} , whose values extend to $+400 \text{ mV}$. The equation is inappropriate for cytochrome oxidase since its haem moiety, haem A (ref. 15), has a larger surface area due to the presence of a farnesyl side chain. A very modest increase in the slope of the line shown in Fig. 2 would raise the maximum E'_0 value to $+400 \text{ mV}$ and thus accommodate the remaining measured values. Unfortunately, the E'_0 value for a monomeric haem in aqueous solution cannot be measured experimentally since haems avidly polymerise by base stacking and/or micelle formation. The E'_0 value for the monomeric haem octapeptide of cytochrome c has been reported¹⁶ to be -217 mV . The equation predicts that about 37% of the surface of this haem would be buried by the peptide atoms. While the structure of the haem octapeptide is unknown, a CPK model indicates that the predicted burial is feasible. Thus, the exposure of the haem moiety, rather than the apolarity of the haem environment or the ligands in the fifth and sixth haem coordination position seems to be the principal determinant for the magnitude of the E'_0 value.

This investigation was supported by a USPHS grant GM-13215 from the Institute of General Medical Sciences. I thank Arthur Arnone, Patrick Briley and Charles Swenson for useful discussions.

EARLE STELLWAGEN

Department of Biochemistry,
The University of Iowa,
Iowa City, Iowa 52242

Received 5 May; accepted 10 July 1978.

- Kassner, R. J., *Proc. natn. Acad. Sci. U.S.A.* **69**, 2263–2167 (1972); *J. Am. chem. Soc.* **95**, 2674–2677 (1973).
- Shrake, A. & Rupley, J. A. *J. molec. Biol.* **79**, 351–371 (1973).
- Teller, D. C. *Nature* **260**, 729–731 (1976).
- Lee, B. & Richards, F. M. *J. molec. Biol.* **55**, 379–400 (1971).
- Chothia, C. *J. molec. Biol.* **105**, 1–14 (1976).
- Horio, T. & Kamen, M. D. *Biochim. biophys. Acta* **48**, 266–286 (1961).
- Hettinger, T. P. & Harbury, H. A. *Biochemistry* **4**, 2585–2589 (1965).
- Scholes, P. B., McLain, G. & Smith, L. *Biochemistry* **10**, 2072–2076 (1971).
- Banerjee, R. & Cassoly, R. *J. molec. Biol.* **42**, 337–349 (1969).
- Cassatt, J. C., Marini, C. P. & Bender, J. W. *Biochemistry* **14**, 5470–5475 (1975).
- Velick, S. F. & Strittmatter, P. *J. biol. Chem.* **221**, 265–275 (1956).
- Cookson, D. J. *et al. Eur. J. Biochem.* **83**, 261–275 (1978).
- Dickerson, R. E. & Timkovich, R. *Enzymes* 3rd edn **11**, 397–547 (1975).
- Henderson, R. W. & Morton, T. C. in *Handbook of Biochemistry*, 2nd edn (ed. Sober, H. A.), J-41–J-49 (CRC Press, Cleveland, 1970).
- Caughy, W. S., Smythe, G. A., O'Keefe, D. H., Maskasky, J. & Smith, M. L. *J. biol. Chem.* **250**, 7602 (1975).
- Harbury, H. A. & Loach, P. A. *J. biol. Chem.* **235**, 2646–3653 (1960).

reviews

Prehistoric standing stone sites

E. MacKie

Sun, Moon and Standing Stones. By J. E. Wood. Pp. 217. (Oxford University Press: Oxford, 1978.) £6.95.

FOR some time now there has been a clear need for a book which presents the findings of the 'archaeoastronomers'—and this means mainly Professor Alexander Thom and his family—on the British prehistoric standing stone sites to the informed general public (an audience which, in this particular case, includes much of the archaeological profession). Thom's discoveries are among the most important and controversial to have confronted prehistorians for many decades; they concern the intellectual qualities which, he claims, lie behind the design and positioning of the stone circles and standing stones and are of a kind which, if accepted, must completely alter our picture of Neolithic British society. Thom's work is, however, couched in highly technical and mathematical terms and can be hard going even for those reasonably competent numerically—though the rewards for perseverance are great.

Dr Wood has succeeded admirably in his task. Here is a lucid, carefully composed and balanced account of the findings of the archaeoastronomical school which makes even the most difficult of these ideas reasonably plain. After reading chapter 7 the reviewer found himself for the first time—after eleven years of experience of this field—with a fair grasp of Thom's theory of the use of the fan-shaped stone rows. This is that they were ancillary apparatus at some lunar observatories which could be used to determine the moon's extreme monthly declinations by extrapolation from the actual observations.

Dr Wood has many sensible things to say about the fallacy of equating the modern skills needed to unravel traces of ancient geometry and astronomy with the prehistoric skills involved. As he says, twentieth century man easily forgets how simple and effective the techniques of the pre-scientific age can be and tends to assume that the ancients must have been phenomenally wise to have done with Stone Age equipment what we use theodolites and electronic calculators to do. In its extreme form, this false premise leads

to the belief that all powerful extra-terrestrials must have been involved. And a milder form of this naive belief leads even some archaeologists to maintain that Thom's deductions are simply not credible because Neolithic man could not have possessed the advanced mathematical and astronomical knowledge that he possesses. Of course they did not, but the same results could still have been achieved from a much simpler technological and intellectual base.

Particularly useful is Dr Wood's careful description in chapter 3 of the methods by which the various geometrical shapes Thom believes were incorporated in stone circle designs—true circles, ellipses, flattened circles and egg-shapes—could have been laid out on the ground using only simple equipment like loops of rope, pegs and measuring rods. This exercise provides a further illustration of why one should not attribute too much abstract knowledge to the megalith builders.

There is little doubt that right-angled triangles were the basis of many of these geometrical shapes and it is easy to conclude that the attraction to the circle-builders of the perfect ones found—the 3:4:5, 5:12:13 and 8:15:17 triangles—was the relationship between the squares on their sides thought to have been discovered by Pythagoras in the sixth century BC.

The idea that the Neolithic circle-builders of Britain anticipated Pythagoras by two millennia or more is compelling and would, if true, considerably alter our ideas on the history of intellectual thought. Dr Wood points out, however, that the fact that some not quite perfect triangles were used—such as $8^2 + 9^2 \approx 12^2$ (actually 12.04)—almost certainly means that these numbers were used because they gave 90° angles when laid out on the ground. We certainly cannot assume that the circle-builders also knew the Pythagorean theorem.

Dr Wood brings the same lucidity and common sense to his review of the astronomical qualities claimed to exist in the standing stone sites and gives us, for example, a useful and up-to-date account of the various theories current about how Stonehenge might have been used as an observatory. Anyone wanting a reliable guide through what sometimes seems like a

maze of theory and speculation about this site cannot do better than this.

Prehistorians will not be satisfied with the relatively meagre amount of space given to the more traditional kinds of archaeological evidence and some will feel that their objections to the ideas of the archaeoastronomical school are not adequately covered (though it is true that little of this is actually in print).

Dr Wood is clearly much more at home with the technical and mathematical evidence than with the historical and archaeological. This shows, for example, in his discussion of the concept of the prehistoric unit of length termed the Megalithic Yard; the statistical uncertainties sway him against the idea but the parallel, historical evidence published by the reviewer—which provides striking independent support for the existence of this unit—is not mentioned. However, the book is intended primarily to be a guide through the complexities of archaeoastronomy and archaeogeometry and it succeeds extraordinarily well. □

Euan MacKie is at the Hunterian Museum, University of Glasgow, UK.

Catastrophe theory

Catastrophe Theory and Its Applications: Surveys and Reference Works in Mathematics. By T. Poston and I. Stewart. Pp.491. (Pitman: London, 1978.) £25.

HERE at last is the first connected account of catastrophe theory written for people who are not pure mathematicians but who want to understand the applications of this new branch of geometry. Previously there has been only Thom's classic *Structural Stability and Morphogenesis*, often as deeply obscure as it is stimulating, and Zeeman's *Collected Papers 1972-1977*, often brilliantly clear but necessarily disconnected.

The first half of Poston and Stewart's book presents the mathematics of catastrophe theory. They explain clearly Thom's listing of the ways in which critical points of functions (that

is, maxima, minima and saddles) of n variables can coalesce as k parameters vary. The central classification theorem is not proved, but the main ingredients of the proof (transversality, stability, and so on) are very fully described, to the point at which it becomes clear why the catastrophe hierarchy depends strongly on k but hardly at all on n . In applications one often encounters functions that are not in the form of one of Thom's standard polynomials; then intricate rules determine which catastrophe is involved. It is one of the pleasant features of this book that a whole chapter is devoted to these. The interplay between algebraic and geometric ideas is emphasised throughout with the aid of a large number of diagrams. Functions are considered in the modern way, as mappings between sets, and the corresponding notation is used abundantly. However, it is explained at the beginning, and there is sufficient reversion to the old-fashioned " $f(x)$ " for traditionally educated scientists not to lose the thread of the argument.

The second half of the book describes some applications. These are

mostly in the physical sciences, so that it is possible to state definitely what 'potential function' it is whose coalescing critical points give the catastrophes. In Zeeman's pretty theory of ship stability, for example, the potential is gravitational energy; in optical caustics it is the ray transit time (more generally, the action); in beam buckling problems it is elastic energy; in fluid flow it is the stream function; and in phase transitions it is the Helmholtz free energy. Usually the fold and cusp yield little not already known to specialists in these various fields; only with the swallowtail, umbilics and higher catastrophes does one start to generate surprising richness of behaviour and genuine applications to the theory as opposed to mere illustrations.

These authors are masters of advanced exposition, and they are to be congratulated on producing what I am sure will become a classic in its subject. The publishers deserve credit for the high standard of presentation, but not for the price.

Michael Berry

Michael Berry is Reader in Physics at the University of Bristol, UK.

appointing, however, in that it deals almost exclusively with electrophoretic methods. Nevertheless, this particular method is described very well. It is less fortunate that, even though the measurement of dielectric constants and dielectric dispersion data feature strongly in the book, they are tucked away in earlier theory chapters rather than in the experimental chapter. Other novel methods such as ion exchange chromatography, isoelectric focusing and precipitation get but brief mentions, and electro-optic phenomena are by-passed altogether.

In conclusion, the topics with which the book deals are treated well. The book provides a very readable and concise account of the electrical and interaction phenomena encountered in biopolymer solutions and makes a valiant attempt to link the fundamentals of electrostatics and biochemistry. However, it is perhaps a pity that, after all the theoretical and conceptual material, the author did not give examples of how the experimental data could be analysed, interpreted and related to the basic properties and functions of macromolecules. A very good handbook for graduate students and researchers in the field of molecular biochemistry in general and for those engaged in electrical or dielectric measurements in particular.

B. Jennings

B. Jennings is Professor of Experimental Physics at Brunel University, Uxbridge, UK.

Molecular biochemistry

Electrical Interactions in Molecular Biophysics. By R. Gabler. Pp. 352. (Academic: New York, San Francisco and London; 1978.) \$25; £16.25.

THIS text forms a very readable handbook which outlines the fundamentals of electrostatic and dielectric theory on the one hand and of chemical bonding and molecular biochemistry on the other. From the beginning, the chemistry is given in terms of organic macromolecules. The author stresses the electrostatic nature of molecular interactions and takes care to develop the concepts of dipole moments and dielectric constants.

A good chapter uses these foundations for a detailed account of the various types of charge and dipolar interactions encountered in real molecules together with some idea of their relative importance. Even though the author uses tables to full advantage in attempts to give the reader a feel for the magnitudes of dipole moments, dielectric constants and polarisabilities in real bonds and molecules, the book lacks the ability to convey confidence and insight as to how the reader should treat, evaluate and really understand the complicated behaviour of say the nucleic acids, globular proteins or poly-

saccharides, for example. Three good chapters form condensed, readable but detailed resumés of Van der Waals forces, Debye-Hückel theory and the structure of water. Emphasis for these is again placed on biopolymer materials.

The final chapter on experimental electrical techniques is somewhat dis-

British bryophytes

The Moss Flora of Britain and Ireland. By A. J. E. Smith. Pp. 706. (Cambridge University Press: Cambridge and London, 1978.) £27.50.

FIFTY-FOUR years have elapsed since the publication of a comprehensive account of British bryophytes. Smith's new book is a worthy replacement for Dixon's *Handbook*, as it combines succinct detail of text with splendid illustrations. It is easier to use than the old *Handbook*, for illustrations are scattered throughout the text rather than being grouped at the end. In many respects it matches the more recent and most popular guide to mosses by E. V. Watson (*British Mosses and Liverworts*; Cambridge University Press second edition, 1968) but surpasses this book in its comprehensive coverage; 692 species are described

and illustrated. It lacks the habit drawings, which were a valuable feature of Watson's book, especially for beginners, but this is not a serious deficiency.

Profuse taxonomic revisions have occurred since Dixon's *Handbook* and the new names will undoubtedly find more widespread acceptance as a result of this publication. *Mnium* now proliferates into *Plagiomnium* and *Rhizomnium*. *Acrocladium* is now *Calliergon* (though we are spared *Calliergonella cuspidata*), and so on. The treatment of that awkward genus *Sphagnum* (contributed by M. O. Hill) is excellent.

This is a book which will be greeted enthusiastically by amateur and professional botanists alike. The price is high but this should not deter the enthusiast; it is worth every penny.

Peter D. Moore

Peter D. Moore is Senior Lecturer in Plant Sciences at King's College, University of London, UK.

Potato research

The Potato Crop: The Scientific Basis for Improvement. Edited by P. M. Harris. Pp. 730. (Chapman and Hall: London; 1978.) £25.

THIS book is one of the first of a series of crop monographs; those on barley and potatoes have appeared and one on wheat is promised. According to the general editor of the series, E. H. Roberts, the aim is to provide authoritative and comprehensive summaries of current scientific knowledge aimed at teachers, students and research and advisory workers (they are not meant as practical manuals for farmers). The project as a whole is timely because the past 20 years has seen a formidable growth of knowledge of the scientific aspects of all our crops, yet there are few good recent monographs of temperate crops available; tropical crops, perhaps surprisingly, have been, on the whole, rather better served.

To what extent, then, does the present work fulfil these intentions? In general, I think it does so very

well. The book, a large one, is composed of 17 chapters which start with history and biosystematics, and go on through morphology, growth physiology, mineral nutrition, water relations, plant density and 'seed' management to mechanisation and weed control; there follow treatments of diseases, pests, tuber quality, storage, breeding, and economics; and the book concludes with a somewhat more speculative (and very interesting) chapter on the application of physiological ideas to the enhancement of crop performance.

The editor has done a good job of making a reasonably coherent whole out of diverse material contributed by 21 authors. The book is well made, the illustrations and index are adequate, and bibliographies (chapter-by-chapter) seem to be fairly complete; misprints and errors are few. Individually, the writers clearly know their subjects well, so the work as a whole is authoritative as well as comprehensive. The facts that our knowledge has improved greatly in recent decades and that the management of the crop and the product is shifting rather rapidly as new technology emerges become very apparent on reading this book. That knowledge and management are both yet imperfect, however, is also clear

and several authors draw attention to deficiencies in these respects.

This book will surely become the standard reference work for the crop and anyone seriously concerned with potato research will need to have it near to hand. Teachers will find it useful and so, perhaps, will some potato farmers (who tend to be knowledgeable enthusiasts); but it is not a practical manual.

There are two defects which should be considered in preparing a revision. First, few readers will find 30 odd pages of taxonomy of wild species very useful in the crop physiological context of this book, and the first two chapters could therefore be condensed into one without loss of content but with some saving of space. Second, though the bibliographies are reasonably international in character, the more agricultural parts of the text are much coloured by Western European, especially UK, practice. A wider viewpoint (including the place and potential of the crop at low latitudes) would be desirable.

N. W. Simmonds

N. W. Simmonds is Lecturer in the School of Agriculture at the University of Edinburgh, UK.

Quantum electrodynamics textbook

Quantum Electrodynamics. By S. N. Gupta. Pp. 226. (Gordon and Breach: New York, Paris and London, 1978.) £25.30.

QUANTUM electrodynamics is the relativistic quantum theory of photons interacting with electrons and other charged leptons. It grew between about 1930 and 1950, as the natural result of combining special relativity with quantum theory. It has made some spectacularly accurate predictions, and it is one of the triumphs of mid-twentieth century physics.

There are few textbooks on just this subject, and the appearance of a new one is noteworthy, particularly when it is by someone like Professor Gupta, who has been a frequent and important contributor to the subject.

One of the best things about this book is its conciseness. Everything is done as economically and elegantly as possible, with no superfluous words. To achieve this compression some things inevitably have had to be sacrificed. I missed any discussion of polarisation,

of bound states, and, most seriously, of the experimental confirmation of the subject. There is a paucity of references, and there are none at all to other textbooks.

On technical matters, the gauge condition is dealt with using the indefinite metric, and regularisation is treated in terms of auxiliary fields. Scattering theory is treated rather perfunctorily. The difficult subject of renormalisation is briefly but, to my taste, sensibly and adequately covered.

Quantum electrodynamics is not a self-contained subject. In calculating the magnetic moment of the electron, for example, weak interactions and the virtual production of strongly interacting particles would have to be included if sufficient accuracy were required. Professor Gupta's book may be one of the last to confine itself to electrodynamics. It is none the worse for that; but a discussion would have been welcome of the place of quantum electrodynamics within modern physics.

This brief and clear book would be a good one from which to learn the subject. It is a pity that the price is too high for most students to pay.

J. C. Taylor

J. C. Taylor is Reader in Theoretical Physics at the University of Oxford, UK.

SCIENTIFIC — BOOKS

H. K. LEWIS can supply works in all branches of Pure and Applied Science. Catalogues on request. Please state interests.

SCIENTIFIC LENDING LIBRARY

Annual Subscription from £7.00. (Available in U.K. only)

Reduced rates for multiple subscriptions.

Prospectus post free on request.

Quarterly List of New Books and new editions added to the Library sent post free to subscribers regularly.

H. K. LEWIS & Co. Ltd.

136 GOWER STREET,
LONDON, WC1E 6BS

Telephone: 01-387 4282
Telegrams: "Publicvit,
London, WC1E 6BS."

Circle No. 30 on Reader Enquiry Card.

obituary

R. G. W. Norrish, 1897-1978

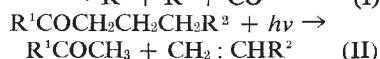
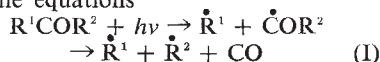
THE death of Professor Ronald George Wreyford Norrish in his 81st year at Cambridge on 7 June 1978 marks the passing of an era notable for remarkable developments in chemistry, among which must be included the advance to maturity of reaction kinetics as a discipline in physical chemistry. Throughout Norrish's research career, extending over more than forty years, he was intimately concerned with a wide variety of subjects, including photochemistry, atomic and free-radical reactions, chain processes such as ignition and combustion and polymerisation, and energy transfer between and within molecules. In all these his studies led to important, and sometimes revolutionary, advances. Wide as his research activities were, yet they were intertwined, with the thread of an intense interest in the behaviour of activated species and the study of the latter by kinetic techniques, running through them.

Norrish's first photochemical work, carried out in 1923 in collaboration with E. K. Rideal, was a study of the photochemistry of aqueous potassium permanganate solutions. At that time photochemistry was in its infancy. The first application of quantum theory to photochemistry had been made by Einstein in 1912 in the form of his law of photochemical equivalence (although in some respects this was anticipated by work of Stark) and its precise applicability to photochemical reactions was still a matter for debate. The remarkable paper by Franck on the elementary processes of photochemical reactions did not appear until 1925. Norrish published a number of interesting papers before 1930, notably those on the photochemistry of nitrogen dioxide, in which he demonstrated a photochemical threshold and the achievement of a photostationary state.

In the period up to the middle 1930s correlation between photodecomposition and photophysical phenomena, such as spectral character and the appearance of fluorescence, engaged the attention of photochemists and Norrish's laboratory in Cambridge occupied a prominent position in these studies. Of his photochemical work before World War II, his long series of investigations of the photolysis of carbonyl compounds is perhaps particu-

larly far-reaching.

The first paper, in 1928, concerned with the photodecomposition of glyoxal, was followed by reports of studies of formaldehyde, acetaldehyde, acetone and a variety of aldehydes and ketones carrying longer hydrocarbon chains. These led to the recognition of the now famous Norrish Type I and Norrish Type II reactions, illustrated by the equations



in which the groups R^1 and R^2 are understood to contain one or two carbon atoms. The observation that Types I and II reactions are influenced strongly by the physical environment—i.e. whether gas phase or condensed phase—provided a most important 'spin-off', namely the first experimental illustration of the Franck-Rabinowitch cage mechanism operative in solution.

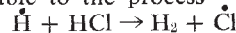
Norrish's interest in intense light sources for photolysis seems to have arisen during World War II. I recall a comment (suitably embellished), made near the end of the war, that he intended to get a searchlight to photolyse keten and to trap the resulting methylene by the Paneth method. At that time, rudimentary ideas of flash photolysis may have been growing in his mind.

Flash photolysis, pioneered by Norrish and Porter and their coworkers, was a truly revolutionary development. It makes use of a powerful short-lived 'photoflash' which brings about rapid dissociation of the absorbing species into radicals or atoms. A second weaker spectroscopic flash ('specflash') triggered electronically at a precise interval after the first, permits the observation of transient species and by varying the time between the two flashes the kinetics of the formation and decay of such species may be studied. A fascinating account of these developments, extending from pyrolysis and combustion processes to vibrational relaxation phenomena, was given by Norrish in his Faraday Lecture to the Chemical Society. The immense impact of this technique on the study of fast reactions and the behaviour of active species is now too well-known to require description and the award of the

Sorry, for copyright reasons some images on this page may not be available online

Nobel Prize for Chemistry in 1967 jointly to Norrish and Porter and to Eigen was universally welcomed.

Norrish's interest in chain processes was evident from his early work on the hydrogen-chlorine reaction in which he elucidated the inhibitory effects of oxygen and ammonia. Adoption of a new technique (continuous photometric analysis) enabled Norrish and Ritchie to establish that the retarding effect of HCl in pure systems and systems containing oxygen is attributable to the process



Norrish's work on the kinetics of this chain reaction, many features of which had defied solution by earlier workers, established his eminent position among kineticists.

Much of Norrish's work on combustion was centred on the transition between slow reaction and ignition and particularly the dependence of the ignition boundaries on the concentrations of various additives (e.g. NO_2 , NOCl , CCl_3NO_2). During the early stages, Semenov was developing his classical theories of thermal and isothermal ignition; Norrish saw the necessity for combining both these in a branched-chain thermal theory. Although some of the detailed interpretations have been modified as the result of applications of new experimental techniques, there is no doubt that the beautiful work of Norrish and his colleagues, made over a long period, has provided much of the basis of our present understanding of these complex processes. The experimental demonstration of degenerate (or delayed) branching in hydrocarbon combustion was an outstanding contribution.

© Godfrey Argent

In the field of polymerisation, Norrish also made pioneering kinetic studies. The best-known is probably that concerned with the acceleration which occurs in the later stages of free-radical polymerisation, to which the name 'gel-effect' has now been attached. Norrish and Smith showed that this is the result of a steadily decreasing rate of chain termination which sets in when the viscosity becomes high. A nice example of the cross-fertilisation of his ideas is provided by his work on the photolysis of poly(methylvinylketone), which was found to undergo photolysis in solution according to both Types I and II. The former yields macroradicals which, in the presence of a polymerisable monomer, give block copolymers.

Norrish had intense energy and enthusiasm and an all-consuming desire to get to the bottom of things by experimental means. These qualities could be disconcerting to his research students, who were sometimes expected to obtain definitive results between 9.30 a.m. and 6.0 p.m. on the same day (bank-holidays included!). Norrish was a generous and warm-hearted, yet forthright, person. His membership of the Savage Club reflected his enjoyment of meeting people, especially those distinguished in professions remote from his own. He had a tremendous zest for life and his hospitality became legendary, not only on the more formal occasions, but also in his entertainment of friends and colleagues from all parts of the world at his home in Park Terrace. There, one might be regaled with recordings of Japanese music or a lecture by Rutherford, or enter into quasi-philosophical discussions of the 'fourth law' of thermodynamics and when Norrish received his honorary degree in the University of Liverpool, the Public Orator recalled the remarkable aptness of Praed's words

'Beginning with the laws which keep
The planets in their radiant courses,
And ending with some precept deep
For dressing eels, or shoeing horses.'
It might truly be said that with Norrish every evening was Christmas eve.

Things Cantabrigian occupied a very special place in Norrish's affections. He was born and educated in Cambridge and effectively spent all his working life there. In 1937, after the death of T. M. Lowry, he had become the second holder of the Chair of Physical Chemistry and subsequently the head of the separate Department of Physical Chemistry. He was justifiably proud of the department he had built up and had strong associations with his College (Emmanuel) of which he was Senior Fellow at the time of his death. His lectures on physical chemistry were

often vividly illustrated by spectacular demonstrations of his favourite phenomena in photochemistry and chain processes. During the war, his department was largely concerned with problems related to the war-effort, such as suppression of gun flash and the development of incendiary materials; during this time Norrish was Chairman of the Incendiary Projectiles Committee.

Norrish's distinction naturally led to his receiving many honours. He was elected Fellow of the Royal Society in 1937 and received honorary doctorates from the Sorbonne, and the Universities of Lancaster, Leeds, Liverpool, Sheffield and British Columbia. He was awarded the Meldola Medal of the Institute of Chemistry (1928), the Davy Medal of the Royal Society and the Liversidge Medal of the Chemical Society (1958), the Lewis Gold Medal of the Combustion Institute (1964), the Faraday and Longstaff Medals of the Chemical Society (1964, 1968) and was a member of eight foreign Academies of Science. During 1953-55 he was President of the Faraday Society, and in 1961 was President of Section B of the British Association for the Advancement of Science. In 1966 he gave the Bakerian Lecture of the Royal Society. Notwithstanding all his honours, he retained an innate modesty which allowed him to accord to others, particularly his junior colleagues, due appreciation of their efforts.

In 1926 he married Anne, the eldest daughter of A. E. Smith, of Manchester, who survives him. They had twin daughters.

C. H. Bamford

J. P. Kendall

PROFESSOR JAMES PICKERING KENDALL, F.R.S., F.R.S.E., Emeritus Professor of Chemistry, University of Edinburgh, died in Edinburgh on 14 June 1978.

He was born on 30 July 1889 in Chobham, Surrey, and was educated at Farnham Grammar School and Edinburgh University. A distinguished undergraduate career was followed by work on exact conductivity measurements, first in Edinburgh and later with Svante Arrhenius at the Nobel Institute, Stockholm. Kendall was inspired by Arrhenius, but found research facilities in the Institute disappointing and after a spell he went to the better equipped Technological Institute of St. Petersburg.

In 1913 Kendall went to Columbia University, New York, as an instructor in chemistry and in 1922 was promoted to a full professorship. He carried out research on a variety of topics includ-

ing the formation of addition compounds in solution and ionisation equilibria. A good example of his incisive and discriminating approach to research is contained in a long, critical and polemical paper on the properties of strong electrolytes in solution in which facts and arguments are marshalled against the views and postulates of Ghosh.

Probably his most important work was the attempt to separate isotopes and mixtures of rare earths, a pioneer effort at a period when the sophisticated armoury of the modern chemist was not available. Isotopic separation eluded him, but by the ionic migration method he effected a substantially complete separation of yttrium from erbium. At the same time he undertook the revision of Alexander Smith's famous series of chemistry textbooks, a task which made increasing demands on his time and energy.

In 1927 Kendall was elected a Fellow of the Royal Society and a year later succeeded Sir James Walker as Professor of Chemistry in the University of Edinburgh, a post he held until his retirement in 1958. He immediately made full use of his administrative ability and business acumen into extending the research facilities of his department, and with the help of men like J. A. V. Butler, E. B. Ludlam, M. Ritchie and H. W. Melville, succeeded in building up a thriving school of physical chemistry. Apart from unsuccessful attempts to separate the hydrogen isotopes in water and to detect the calcium isotope with mass 41 in the potassium-rich deposits of Rhiconich in Sutherlandshire Kendall gave up research and concentrated on writing books. These included *At Home among the Atoms and Young Chemists and Great Discoveries* and the lively and readable biographies of Michael Faraday and Humphry Davy. In 1938 he wrote *Breathe Freely* in which he correctly predicted that in the event of war poison gas would not be used.

Kendall was an excellent lecturer and he gave the 1938/39 course of Christmas lectures at the Royal Institution, in London. In one of the lectures he performed the famous boiling lead experiment, when his daughter Jean passed her hand through the stream of white hot metal. He also took a great interest in the Royal Society of Edinburgh and served first as General Secretary (1936-1946) and then as President (1949-1954). In his younger days he enjoyed bicycling and he was a very good bridge player. He was twice married and his second wife, Jean, tended him devotedly in his latter years of ill-health. He leaves a son and two daughters.

Neil Campbell

Recent scientific and technical books

Biology

SMITH, John Maynard. *The Evolution of Sex*. Pp.x+222. ISBN-0-521-21887-X. (Cambridge, London and New York: Cambridge University Press, 1978.) Hard cover £9.50; Paperback £2.95.

STILES, W. S. *Mechanisms of Colour Vision*. (Selected papers of W. S. Stiles, F.R.S., with a new introductory essay.) Pp.298. ISBN-0-12-671350-2. (London and New York: Academic Press, a Subsidiary of Harcourt Brace Jovanovich, Publishers, 1978.) £14; \$27.35.

WHITTAKER, Peter A., and DANKS, Susan M. *Mitochondria: Structure, Function and Assembly*. (Integrated Themes in Biology.) Pp.xii+145. ISBN-0-582-44382-2. (London and New York: Longman Group, Ltd., 1978.) £3.75.

Applied Biology

ALVIM, Paulo de T., and KOZLOWSKI, T. T. (edited by). *Ecophysiology of Tropical Crops*. Pp.xiv+302. ISBN-0-12-055650-2. (New York and London: Academic Press, a Subsidiary of Harcourt Brace Jovanovich, Publishers, 1977.) \$29.50.

ANFINSEN, C. B., EDSALL, John T., and RICHARDS, Frederic M. (edited by). *Advances in Protein Chemistry*, Vol. 31. Pp.vii+449. ISBN-0-12 034231-6. (New York and London: Academic Press, a Subsidiary of Harcourt Brace Jovanovich, Publishers, 1977.) \$31.

AUTOR, Anne Pomeroy (edited by). *Biochemical Mechanisms of Paraquat Toxicity*. Pp.x+240. ISBN-0-12-068850-6. (New York and London: Academic Press, a Subsidiary of Harcourt Brace Jovanovich, Publishers, 1977.) \$12.50.

BALDWIN, R. W. (edited by). *Secondary Spread of Cancer*. Pp.244. ISBN-0-12-076850-X. (London and New York: Academic Press, a Subsidiary of Harcourt Brace Jovanovich, Publishers, 1978.) £12.40.

BLACKMAN, D. E., and SANGER, D. J. (edited by). *Conemporary Research in Behavioral Pharmacology*. Pp.xiv+506. ISBN-0-306-31061-9. (New York and London: Plenum Press, 1978.) \$47.40.

BODMER, W. F., BATCHELOR, J. R., BODMER, J. G., FESTENSTEIN, H., and MORRIS, P. J. (edited by). *Histocompatibility Testing 1977*. (Report of the 7th International Histocompatibility Workshop and Conference, Oxford, England, 4-10 September 1977.) Pp.612. ISBN-87-16-02735-3. (Copenhagen: Munksgaard, 1978.) D.kr.245.

BRIGGS, D. E. Barley. Pp.xviii+612. ISBN-0-412-11870-X. (London: Chapman and Hall, Ltd., 1978. Distributed in the USA by Halsted Press, a Division of John Wiley and Sons, Inc., New York.) £25.

BRITISH PHARMACOPŌEIA 1973. ADDENDUM 1978. (Published on the recommendation of the Medicines Commission pursuant to the Medicines Act 1968.) Pp.xxii+57+61. ISBN-0-11-320689-5. (London: HMSO, 1978.) £6.50 net.

BUSCH, Harris (edited by). *Methods in Cancer Research*, Vol. 14. Pp.xix+374. ISBN-0-12-147674-X. (New York and London: Academic Press a Subsidiary of Harcourt Brace Jovanovich, Publishers, 1978.) \$37; £31.50.

BUTLER, Francine. *Biofeedback: a Survey of the Literature*. Pp.xi+340. ISBN-0-306-65173-4. (New York, Washington and London: IFI/Plenum, 1978.) \$54.

CERVOS-NAVARO, J., BEITZ, E., EBHARDT, G., FERSZT, R., and WULLENWEBER, R. (edited by). *Pathology of Cerebrospinal Microcirculation*. (Advances in Neurology, Vol. 20.) Pp.xxvi+606. ISBN-0-89004-237-3. (New York: Raven Press, 1978.) \$58.50.

CHICHESTER, C. O., MRAK, E. M., and STEWART, G. F. (edited by). *Advances in Food Research*, Vol. 23. Pp.vii+436. ISBN-0-12-016423-X. (New York and London: Academic Press, a Subsidiary of Harcourt Brace Jovanovich, Publishers, 1977.) \$29.

COTMAN, Carl W. (edited by). *Neuronal Plasticity*. Pp.xiv+335. ISBN-0-89004-210-1. (New York: Raven Press, 1978.) \$32.50.

DAVIS, W., and HARRAP, K. R. (edited by). *Characterization and Treatment of Human Tumours: Proceedings of the Seventh International Symposium on the Biological Characterization of Human Tumours*. Budapest, April 13-15, 1977. (Advances in Tumour Prevention, Detection and Characterization, Vol. 4.) Pp.332. ISBN-0-444-90014-4. (Amsterdam and Oxford: Excerpta Medica, 1978.) Dfl. 115; \$50.

DENIKER, P., RADOUCO-THOMAS, C., and VILLENEUVE, A. (chief editors). *BARONET-LACROIX, D., and GARCIN, F.* (associate editors). *Neuro-Psychopharmacology*. (Proceedings of the Tenth Congress of the Collegium Internationale Neuro-Psychopharmacologicum, Québec, July 4-9, 1976.) Vol. 1: xxvii+1-932. Vol. 2: xix+933-1737. ISBN-0-08-21506-8. (Oxford and New York: Pergamon Press, 1978.) £85.

EVANS, W. Howard. *Preparation and Characterisation of Mammalian Plasma Membranes*. (Laboratory Techniques in Biochemistry and Molecular Biology.) Pp.266. ISBN-0-70204-422-2. (Amsterdam, New York and Oxford: North-Holland Publishing Company, 1978.) Dfl.60; \$26.

FRIGERIO, Alberto (edited by). *Recent Developments in Mass Spectrometry in Biochemistry and Medicine*. Vol. 1. Pp.xii+658. ISBN-0-30631138-0(v.1). (New York and London: Plenum Press, 1978.) \$55.

GEMME, Robert, and WHEELER, Connie Christine (edited by). *Progress in Sexology*. (Selected papers from the Proceedings of the 1976 International Congress of Sexology.) (Perspectives in Sexuality: Behaviour, Research and Therapy.) Pp. xviii+615. ISBN-0-306-31104-6. (New York and London: Plenum Press, 1977.) \$39.

GEORGE, William J., IGNARRO, Louis J. (edited by). *Advances in Cyclic Nucleotide Research*, Vol. 9. (Third International Conference on Cyclic Nucleotides, New Orleans, Louisiana, U.S.A., July 17-22, 1977.) Pp.xxiii+799. ISBN-0-89004-240-3. (New York: Raven Press, 1978.) \$75.40.

GUPTA, I. C., and PAHWA, K. N. *A Century of Soil Salinity Research in India: An Annotated Bibliography 1863-1976*. Pp.xii+400. (New Delhi, Bombay and Calcutta: Oxford and IBH Publishing Co., 1978.) Rs.72.

HANAN, Joe J., HOLLEY, Winifred F., and GOLDSBERRY, Kenneth L. *Greenhouse Management*. (Advanced Series in Agricultural Sciences, Vol.5.) Pp.xiv+530. ISBN-0-540-08478-9. (Berlin and New York: Springer-Verlag, 1978.) DM.94; \$47.

HOLMQUIST, R. (scientific editor). STICKLAND, A. C. (executive editor). *Life Sciences and Space Research*, Vol. 16. (Proceedings of the Open Meetings of the Working Group on Space Biology of the Twentieth Plenary Meeting of COSPAR, Tel Aviv, Israel, 7-18 June 1977.) Pp.ix+157. ISBN-0-08-022022-3. (Oxford and London: Pergamon Press, 1978.) £14; \$25.

JACOBSON, Marcus. *Developmental Neurobiology*. Second edition Pp.xii+562. ISBN-0-306-31076-7. (New York and London: Plenum Press, 1978.) \$45.

JARON, Dov (edited by). *Proceedings of the Sixth New England Bioengineering Conference*, March 23-24, 1978, University of Rhode Island, Kingston, Rhode Island. Pp.xxi+421. ISBN-0-08-022678-7. (Oxford and New York: Pergamon Press, 1978.) \$40.

JIRSCH, D. W. (edited by). *Immunological Engineering*. Pp.x+361. ISBN-0-85200-174-6. (Lancaster: MTP Press, Ltd., 1978.) £12.95.

JONES, Peter W., and FREUDENTHAL, Ralph I. (edited by). *Polynuclear Aromatic Hydrocarbons: Second International Symposium on Analysis, Chemistry, and Biology*. (Carcinogenesis: a Comprehensive Survey, Vol. 3.) Pp.xx+487. ISBN-0-89004-241-1. (New York: Raven Press, 1978.) \$51.35.

Psychology

ANISMAN, Hymie, and BIGNAMI, Giorgio (edited by). *Psychopharmacology of Aversively Motivated Behavior*. Pp.xii+564. ISBN-0-306-31005-4. (New York and London: Plenum Press, 1978.) \$42.

CAMPBELL, Robin N., and SMITH, Philip T. (edited by). *Recent Advances in the Psychology of Language*. Part A: Language Development and Mother-Child Interaction. Pp.xii+485. ISBN-0-306-32884-4. (v.4a). Part B: Formal and Experimental Approaches. Pp.xiii+432. ISBN-0-306-32885-2. (v.4b). (NATO Conference Series, 3: Human Factors, Vol. 4a and 4b.) (New York and London: Plenum Press, 1978. Published in Coordination with NATO Scientific Affairs Division.) \$22.80 each part.

CATTELL, Raymond B. *The Scientific Use of Factor Analysis in Behavioral and Life Sciences*. Pp.xxii+618. ISBN-0-306-30939-4. (New York and London: Plenum Press, 1978.) \$39.

CLIFFORD, Brian R., and BULL, Ray. *The Psychology of Person Identification*. Pp.xiii+254. ISBN-0-7100-8867-1. (London, Henley and Boston, Mass.: Routledge and Kegan Paul, 1978.) £7.95.

COX, Murray. *Structuring the Therapeutic Process: Compromise with Chaos—The Therapist's Response to the Individual and the Group*. (Pergamon International Library of Science, Technology, Engineering and Social Studies.) Pp.xvi+300. ISBN-0-08-020402-3. (Oxford and New York: Pergamon Press, 1978.) Hard £12.50; Flexi £5.

GAZZANIGA, Michael S., and LeDOUX, Joseph E. *The Integrated Mind*. Pp.xi+168. ISBN-0-306-31085 6. (New York and London: Plenum Press, 1978.) \$19.14.

HARDY, William G. *Language, Thought, and Experience: a Tapestry of the Dimensions of Meaning*. Pp.ix+318. ISBN-0-8391-1213-0. (Baltimore, Md.: University Park Press, 1978. Distributed in U.K. by M.T.P. Press, Ltd., Lancaster.) £14.95.

HOLT, Robert R. *Methods in Clinical Psychology*. Vol. 1: Projective Assessment. Pp.xv+344. ISBN-0-306-31053-8 (v.1). Vol. 2: Prediction and Research. Pp.xiv+337. ISBN-0-306-31124-0 (v.2). (New York and London: Plenum Press, 1978.) \$21 each volume.

LEEUWENBERG, E. L. J., and BUFFART, H. F. J. M. (edited by). *Formal Theories of Visual Perception*. Pp.xii+345. ISBN-0-471-99586-X. (Chichester and New York: John Wiley and Sons, 1978.) £15.75.

LEWIS, Michael, and ROSENBLUM, Leonard A. (edited by). *The Development of Affect*. (Genesis of Behavior, Vol. 1.) Pp.xii+426. ISBN-0-306-34341-X. (New York and London: Plenum Press, 1978.) \$27.

MAHER, Brendan A. (edited by). *Contributions to the Psychopathology of Schizophrenia*. (Selected Reprints, with New Commentaries, from "Progress in Experimental Personality Research.") Pp.ix+382. ISBN-0-12-465250-6. (New York and London: Academic Press, a Subsidiary of Harcourt Brace Jovanovich, Publishers, 1977.) \$9.50.

McGILL, Thomas E., DEWSBURY, Donald A., and SACHS, Benjamin D. (edited by). *Sex and Behavior: Status and Prospects*. Pp.xxii+436. ISBN-0-306-31084-8. (New York and London: Plenum Press, 1978.) \$35.40.

SCHWARTZ, Gary E., and SHAPIRO, David. (edited by). *Consciousness and Self-Regulation: Advances in Research and Theory*, Vol. 2. Pp.xiv+445. ISBN-0-306-33602-2. (New York and London: Plenum Press, 1978.) \$27.

WALK, Richard D., and PICK, Jr., Herbert L. (edited by). *Perception and Experience*. (Perception and Perceptual Development: a Critical Review Series, Vol. 1.) Pp.xiv+432. ISBN-0-306-34381-9. (New York and London: Plenum Press, 1978.) \$30.

WELCH, Robert B. *Perceptual Modification: Adapting to Altered Sensory Environments*. (Academic Press Series in Cognition and Perception.) Pp.xv+346. ISBN-0-12-741850-4. (New York and London: Academic Press, a Subsidiary of Harcourt Brace Jovanovich, Publishers, 1978.) \$24.50; £15.90.

Education

CROPLEY, A. J., and DAVE, R. H. *Lifelong Education and the Training of Teachers: Developing a Curriculum for Teacher Education on the Basis of the Principles of Lifelong Education*. (Advances in Lifelong Education.) Pp.xiv+245. ISBN-0-08-023008-3. (Oxford and New York: Pergamon Press; Hamburg: Unesco Institute for Education, 1978.) Hard £12.50; Flexi £5.

PICKERING, Sir George. *Quest for Excellence in Medical Education: a Personal Survey*. Pp.vi+102. ISBN-0-19-721309-5. (Oxford and London: Oxford University Press, 1978. Published for the Nuffield Provincial Hospitals Trust.) £4.50 net.

SKAGER, Rodney. *Lifelong Education and Evaluation Practice: a study on the Development of a Framework for Designing Evaluation Systems at the School Stage in the Perspective of Lifelong Education*. (Advances in Lifelong Education Vol. 4.) Pp.xiv+155. ISBN-0-08-021813-X. (Oxford and New York: Pergamon Press; Hamburg: Unesco Institute for Education, 1978.) Hard £8; Flexi £4.

History of Science

COHEN, I. Bernard. (edited, with a general introduction by). Isaac Newton's Papers and Letters on Natural Philosophy and Related Documents. Second edition. Pp.xvii+540. ISBN-0-674-46853-8. (Cambridge, Mass. and London: Harvard University Press, 1978.) £18.55.

CRAVENS, Hamilton. *The Triumph of Evolution: American Scientists and the Heredity-Environment Controversy 1900-1941*. Pp.xvi+351. ISBN-0-8122-7744-9. (Philadelphia, Pa. and Lizard Town, South of Helston, Cornwall: University of Pennsylvania Press, 1978.) \$17.50.

DEWHURST, Kenneth, and REEVES, Nigel. *Friedrich Schiller: Medicine, Psychology and Literature*. Pp.xii+413. ISBN-0-9501528-3-8. (Oxford: Sandford Publications, 1978. Distributed in USA and Canada by University of California Press.) £12 net.

GEISON, Gerald L. *Michael Foster and the Cambridge School of Physiology: The Scientific Enterprise in Late Victorian Society*. Pp.xix+401. (Princeton N.J. and Guildford: Princeton University Press, 1978.) £18.40.

KING, Lester S. *The Philosophy of Medicine: The Early Eighteenth Century*. Pp.viii+291. ISBN-0-674-66585-6. (Cambridge, Mass. and London: Harvard University Press, 1978.) £12.25.

KROHN, Wolfgang, LAYTON, Jr., Edwin T., and WEINGART, Peter (edited by). *The Dynamics of Science and Technology: Social Values, Technical Norms and Scientific Criteria in the Development of Knowledge*. (Sociology of the Sciences: a Yearbook, 2) Pp.x+293. ISBN-90-277-0880-0. (Dordrecht, Holland and Boston, Mass.: D. Reidel Publishing Company, 1978.) Cloth Dfl.70; \$29.50. Paper Dfl.35; \$14.95.

NASR, Seyyed Hossein. *An Introduction to Islamic Cosmological Doctrines: Conceptions of Nature and Methods Used for Its Study by the Ikhwan Al-Safa Al-Biruni, and Ibn Sina*. Revised edition. Pp.xiii+318. ISBN-0-500-01190-7. (London: Thames and Hudson, 1978.) £10.

PELLING, Margaret. *Cholera, Fever and English Medicine, 1825-1865*. (Oxford Historical Monographs.) Pp.x+342. ISBN-0-19-821872-9. (Oxford and London: Oxford University Press, 1978.) £7.50 net.

SCHWABE, Calvin W. *Cattle, Priests, and Progress in Medicine*. (The Wesley W. Spink Lectures on Comparative Medicine, Vol. 4.) Pp.xiii+277. ISBN-0-8166-0825-3. (Minneapolis: University of Minnesota Press, 1978.) \$19.75.